

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 04:25:46 ; Search time 4250.63 Seconds
(without alignments)
11416.811 Million cell updates/sec

Title: US-09-874-390-1
Perfect score: 2319
Sequence: 1 atcaactggttcctcatt.....ataataataatcaacaga 2319

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenDbml:*

- 1: gb_ba:*
- 2: gb_hcg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_hcg_hum:*
- 31: em_hcg_inv:*
- 32: em_hcg_other:*
- 33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match Length	ID	Description

1	2194.2	94.6	2216	6	AX045249	AX045249 Sequence
2	2183.2	94.1	2217	6	AX045251	AX045251 Sequence
3	2183.2	94.1	2217	6	AX045253	AX045253 Sequence
4	2081	89.7	2106	9	BC017032	BC017032 Homo sapi
5	1990.8	85.8	2128	6	AF136078	AF136078 Sequence
6	1990.8	85.8	2128	9	AF102542	AF102542 Homo sapi
7	1883.8	81.2	187275	2	AC092755	AC092755 Homo sapi
8	1883.8	81.2	187275	6	AX087948	AX087948 Sequence
9	1317	56.8	1317	6	AF038650	AF038650 Homo sapi
10	1011	43.6	108873	14	AF318573	AF318573 Bovine he
11	1001.4	43.2	2017	14	AF231105	AF231105 Bovine he
12	960	41.4	4179	14	BHVAREGF	BHVAREGF Bovine herp
13	923.8	39.8	1841	10	BC018297	BC018297 Mus muscu
14	726.2	31.3	176483	2	AC096432	AC096432 Rattus no
15	376.8	16.2	2033	10	MMU19265	MMU19265 Mus muscu
16	375.2	16.2	2023	10	D87333	D87333 Mus muscu
17	375.2	16.2	2047	10	D87332	D87332 Mus muscu
18	374.2	16.1	377	6	AX341786	AX341786 Sequence
19	369.4	15.9	1287	6	AX087946	AX087946 Sequence
20	369.4	15.9	2105	6	I41262	I41262 Sequence 3
21	369.4	15.9	2105	6	I61440	I61440 Sequence 3
22	369.4	15.9	2105	6	I72548	I72548 Sequence 3
23	369.4	15.9	2110	6	AX333255	AX333255 Sequence
24	369.4	15.9	2110	6	AX335500	AX335500 Sequence
25	369.4	15.9	2110	9	HUMGOGG	HUMGOGG Human beta-
26	369.4	15.9	2204	9	HUMG2CNT	HUMG2CNT Homo sapi
27	367.8	15.9	159930	2	AL138757	AL138757 Homo sapi
28	367.8	15.9	183858	9	AL161626	AL161626 Human DNA
29	364.6	15.7	5010	10	S79797	S79797 enzymatic g
30	364.2	15.7	2435	4	BTU41320	BTU41320 Bos taurus
31	307.2	13.2	309	6	AX341869	AX341869 Sequence
32	305.6	13.2	309	6	AX340693	AX340693 Sequence
33	257	11.1	380	14	HSBROVHEZF	HSBROVHEZF Bovine herp
34	246.4	10.6	174323	2	AC092754	AC092754 Homo sapi
35	243.2	10.5	3083	2	HDMB16NACT	HDMB16NACT Homo sapi
36	243.2	10.5	151136	2	HSAL189K21	HSAL189K21 Homo sapi
37	207.8	9.0	189534	2	AC095187	AC095187 Human DNA
38	203	8.8	252504	2	AC073815	AC073815 Rattus no
39	198.8	8.6	155359	9	AL358787	AL358787 Mus muscu
40	198.8	8.6	155359	9	AL358787	AL358787 Human DNA
41	191.8	8.3	1362	6	AX087935	AX087935 Homo sapi
42	191.8	8.3	3435	9	AF132035	AF132035 Sequence
43	189.6	8.2	3508	10	AB037596	AB037596 Homo sapi
44	188.6	8.1	184590	2	AC093259	AC093259 Homo sapi
45	185	8.0	63853	2	AC087663	AC087663 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AX045249
DEFINITION Sequence 1 from Patent WO0066727.
ACCESSION AX045249
VERSION AX045249.1 GI:11343799
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Adolf.G., Helder.K.H. and Sommergruber.W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 0066727-A 1 09-NOV-2000;
Boehringer Ingelheim International GmbH (DE)
FEATURES
source 1..2216
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..426
427..1743
/note="unnamed protein product"

3'UTR
BASE
COUNT
ORIGIN

600 a 499 c 547 g 570 t

1744..2216
YGAADLNMNNHNLHLLANKPDPKVDNALQCLBEYLKATYGEVL"
TLOARMMGSLVPHNPKRYDIDSMSTAIRLVWQHGGIDKGAAYACSGIHNQALCV
HLTKKKDKPPYNLTFMGNAIVASREYOHVLYLKNPSQOLIEWVDYTSPEEHLNMA
PWKFYFLTSCGDPFIKSNAEWQALMKMGINSMESEVPRKHETRMKYHEVEVRLDT
CVNDEKSPENFEKAUVAIILSCFPNVEIASLVYVYVASRWYADINCMCDLQSSY
YLSTLRDCEHKEAKREKIQPLPESKEVEEFLAYISVMVHKEIENFERLLRAYAPOONY
RESQSQRNTLNYFLKLPARKSINCSCVTRGDDEAVYQALINLNEKKKKEEPTDTIH
/codon_start=1
/protein_id="CAC17334.1"
/db_xref="GI:11343800"
/translation="MVQMKRLCQLHYLMALCTYMLLTALVAKLSFRLKCDSDHIGLES

Query Match	94.6%	Score 2194.2	DB 6	Length 2216
Best Local Similarity	99.9%	Pred. No. 0		
Matches 2196	Conservative	0	Mismatches 3	Indels 0
				Gaps 0

QY	70	gttaagggaagacgacctgtgctaactgtgacccgaccttattcccaagattttgtgtctgg	129
Db	1	gttCACGGGGAACGCCCCCTGTCTACTGTGTGACCTGGCCCCCTTACTCACACAGTTTGTGTGGG	60
QY	130	gaagccctggggaattctgtcaataactataactgtatgtgtctaaagggaacaacagatgaaga	189
Db	61	GAAGCCCTGGGATTCCTGGTAATACCTATACCTGTAGTGTCCTAAGGGAACAAGATGAAAGA	120
QY	190	acaatgaacctcaagagagcttctgtlcaatgatgaagaagcaagaactgacgctgtgcaaatat	249
Db	121	ACATGACCTCAAGGAGGCTCTCTGTCAATGAGAAAGCAAGCTGAGCGCTGGCAAGATAT	180
QY	250	taaagaagagacctgaagaactgtccctgtgacaactataatgaatgtatgaataatccctttgg	309
Db	181	TAAAGAGAGGCTTGAAACTGTTCCTTGGAACATCTTATGATGTCTGAAATAATCCTTTGG	240
QY	310	agggttaagaagatcagagggacatggtltgtlcaacattgtctgcacaaggaacacccgacgc	369
Db	241	AGGGTTAGAAAGATCAGGGGACATGGTTGTTCATTTGCTGGCAGGAAACACGGCAGTC	300
QY	370	ttcaacttggaaacagaatcaacgccttgtgaagaatcalccctaaagcaggagaagata	429
Db	301	TTTCACCTTGGGAACAACAATCACCGCTTGTGAAAGATATCCTTAAGCAGGAAGAAGCTA	360
QY	430	ctaaaggaattgtgtctctctccacattccctgtgtctggtgtcctcaacggtcccccattct	489
Db	361	CTTAAAGATTGTGTACTCTCTCCACTTCCCTGTGTGTGGTCTCCACCTGTCTCCATTCT	420
QY	490	gttcaagatgtgtlcaatgtgaagaagactctgtccagctgtacattactgtgtgctctggagctgc	549
Db	421	GTAGAGATGGTTCAMTGGAGAAGACTGTGGCCAGCTGATTAATTGTGGGCTCTGGGGCTGC	480
QY	550	tatatgtcgtcgtggccaactgtgtgctctgaaactttcttcaagttgtaagtgtacctctgac	609
Db	481	TATATGCTGTGTGGCCACTGTGGCTGTGAAAATTTTTCAGGTTTAAAGTGTACCTGTAC	540
QY	610	caacttgggtctcgtgagttccagggatcctcaaaacagagtaactgtagaataatctgtfataat	669
Db	541	CACTTGGGCTGTGGAATCCAGGGAAATCTCAAAACCAGTACTGTAGAAATATCTTGTATAT	600
QY	670	ttccctgaaaactlccagcaaaagaggtctlcaactgtltcaaggggtlcaaccgaaaggagcaaa	729
Db	601	TTTCTGAAACCTTCCAGCAAAAGAGGTATTCATCACTGTTCAGGGGCTCACCCGAGGGACCA	660
QY	730	gaaggcaatgtcttcaagctatcttgaataaactctgagaggtlcaagaagaagcagagccttctc	789
Db	661	GAGGAGTGCTTCAAGGCTATTCTGATATACTCGGAGGCTCAAAATAAGGAAGCAAGCCTTTC	720
QY	790	acaagcacccactatactctccctcaacaaagagactgtgaagactlcaagctgtaaaggag	849
Db	721	ACAAGACCCCACTACTCTCCCTCTCACAGAGACTGTGAGCACTTCAAGGCTTAAAGGAAG	780
QY	850	ttcatacagtttcccaactgagcaaaagaagagtgtgaagtttccctattgtcatactctatgtgtg	909

[illegible]

Db	1861	ATTAAGAGGGCTGCTATTAGATTGTGGGTAAGATGATCTTTGCCCTTGCAAAATGCTGCT	1920
Qy	1990	gggtgaatgctgtctgtctctcaccctbaacctagatgtctcctcaacttctca	2049
Db	1921	GGGTGAATGCTGCTGTCTCTCCACCCCTAACCCTAGTAGTCTCTCCACTTCTTCA	1980
Qy	2050	ctaaatggaatggaactgtctgtatagggagatgaagagagatatgtgtagaca	2109
Db	1981	CTAAATGGAATGAAGAACTGCTGTATAGGGAGTGAAGGAGGATATGTGTGAGACA	2040
Qy	2110	cttgattcagttgatgacgtcgtgtgtgagctttccattctgttgagcgtccgtccaa	2169
Db	2041	CTTGATTTCAGTTGATGATGCTCTCTGTAGCTTTTCCATTCTGTGAGACTGCCCTTCTTA	2100
Qy	2170	taattccagtttggtagcgttggaggaacttgaatgaagaagaaccttcccttgt	2229
Db	2101	TAAATCCAGGTTGGTAGCGTGGAGAGAACTTGTATGGAAGAGAACCTTCCCTGTGT	2160
Qy	2230	actgttaacttaaaataatagctccctgattcaagta	2268
Db	2161	ACTGTTAACTTAAATAATAGCTCTGATTCATCAAGTA	2199
RESULT	2		
LOCUS	AX045251	2217 bp	DNA
DEFINITION	Sequence 3 from Patent WO0066727.		linear
ACCESSION	AX045251		
VERSION	AX045251.1	GI:11343801	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2217)		
TITLE	Adolf, G., Heider, K.H. and Sommergruber, W.		
JOURNAL	Tumour-associated antigen		
FEATURES	Patent: WO 0066727-A 3 09-NOV-2000; Boehringer Ingelheim International GmbH (DE)		
source	location/Qualifiers		
	1..2217		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
5'UTR	1..844		
5'UTR	1..426		
CDS	427..999		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC17335.1"		
	/db_xref="GI:11343802"		
	/translation="MVMKRLQQLHYLWALGCVMLATVALKLSFRLKCDSDHGLLES		
	RESGQRCNLTLYNFKLIPAKMSINGSVTRGDEAVLOALINLLEKKREPTDTH		
	YLSTPCEHKKARKEIQLPPLSKSEVEFPIAYSMVHEKIENERRLLRAYAPONTY		
	CVNDEVEYPRNFQSGSNFLLPKLHSD		
	845..1744		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC17336.1"		
	/db_xref="GI:11343803"		
	/translation="MRLKTLKGYCELCMLPRLTVSMMKKSPETEKEAVKAIISCF		
	PNVFISKLAVVYASWSRVQADLCMEDLIQSSVPKVFINTCGDIFPKISNEMVO		
	ALKMLNRSNISEVPPKHKETRWKHYHEFVRLDLHLTKKDKDPSPVNLMEFTNAYI		
	VASDPEFQVHLKPKKSOQLTEWVKDYSPEHLMATLQRAKMPGCSVPNPKDISIM		
	TSIARLYKMGHGEDIDKGPAYPCSSIHQRAICVYAGDLNMLQNHHLANKFDPK		
	VDDNALQCLEBYLTKAIYTEL		
	1000..2217		
3'UTR	1000..2217		
BASE COUNT	600 a 498 c 547 g 572 t		
ORIGIN			
Query Match	94.1%; Score 2183.2; DB 6; Length 2217;		
Best Local Similarity	99.8%; Pred. No. 0;		

QY 1149 gccctggaatactctcctgaatacattgtaggacggaacttccctataaagaagcaatgcaga 1208
 |||||||
 Db 1081 GCCGTGGAATACTCTCTGTAATGATGAGAGGAGCACTTCTTATAAGAGCAATGAGAGA 1140
 QY 1209 gatggtccaggcctcctaagaatgttgaatgtaggaagaatagaatgagatcagaagtactcc 1268
 |||||||
 Db 1141 GATGGTCCAGGCTCTCAAGATGTTGAATGGAGGAATAGCATGAGTCAAGAGTACTCTCC 1200
 QY 1269 taagcaacaagaaccgcctggaataatacactttagagtagtagagacacatcacct 1328
 |||||||
 Db 1201 TAAGCACAAAGAAACCCGCTGGAAATATCATTTGAAGTAGTAGAGACACATTACACT 1260
 QY 1329 aaccaacaagaagaagatccctccctataatctaatactatgttcaagggaatgcgtta 1388
 |||||||
 Db 1261 AACCAACAAGAAAGAGATCTCCCTTATTAATTAATGATGTTACAGGAATGCGTA 1320
 QY 1389 catgtggtctcccgagatctgtccaacatgttttgaagaaccctaataccaacaact 1448
 |||||||
 Db 1321 CATTTGGGCTTCCCGAGATTTGCTCCAAATGTTTTCMAAACCCCTAAATCCCAACAACT 1380
 QY 1449 gatggaatgggtaaagaacacttatagcccaagatgaacacctctgggccaaccttgaag 1508
 |||||||
 Db 1381 GATTGAATGGGTAAAGACACTTATAGCCCAATGAACACCTCTGGGCCACCTTCAGCG 1440
 QY 1509 tgcacggtggaatgcctgctctgttcccaacacaccccaagtaacacatcacaacatgac 1568
 |||||||
 Db 1441 TGCACGGTGAATGCTGCTGCTGTCCCAACCAACCCCAAGTACAGCATCTCAGACATGAC 1500
 QY 1569 ttctattgccaagctgttcaagtaggcagaggtcatagaggagacatcgataaggtgtctc 1628
 |||||||
 Db 1501 TTCTATTGTCAGGCTGGTCAAGTGCAGGGGTCAATGAGGAGACATCATATAGGGTGTCTCC 1560
 QY 1629 ttagtcacctgtctctggaataccacagcggtatctgcttataatgaggctggagact 1688
 |||||||
 Db 1561 TTATGCTCCTGCTCTGGAATCCACAGCGGCGCTATCTGCGTTATAGGGCTGGGCACTT 1620
 QY 1689 gaattggaatgcttaaaacacatcactgttggccaacaagtctgaaccaaggtaga 1748
 |||||||
 Db 1621 GAATTGGATGCTTCAAAACCATCACTGTTGGCAACAAAGTTGACCACAAAGTAGATGA 1680
 QY 1749 taatgctctcaatgcttaaaagaatacctcaatgaatgaagccatctatagtgactgaact 1808
 |||||||
 Db 1681 TAATGCTCTTACGCTTAAAGAAATACCTAATGATTAAGAGCCATCTATGAGCTGACTGACT 1740
 QY 1809 ttgagacacataagaaaggttgcaccctgtgggcaagagcaatgatacaacaatgctca 1868
 |||||||
 Db 1741 TTGAGACACACTATGAGAGCGTTGCTACTCTGTGGGGCAAGAGCATATACAAATGCTCA 1800
 QY 1869 gaactgctggaagaatgtggttgggaagaccagggcttgcgaatctgtggtcatctttag 1928
 |||||||
 Db 1801 GAACCTGCTGGGAGAGTGTGGGGAGACAGGCGCTTCAATTCGTGGCATCTCTTAG 1860
 QY 1929 gataaagagagctctatagattgttgggtgaatagatccttttgccttggaattgtgcgc 1988
 |||||||
 Db 1861 GATTAAGAGGCGCTGCTAATTAATTTGGGTAAAGATCTTTTGCTTTGCAAAATTTGCTGCC 1920
 QY 1989 tgggtgaatgctgttcttcaacccctaaccctagtagtctccctacataactcttc 2048
 |||||||
 Db 1921 TGGGTGAATGCTGCTGTTGCTCTACCCCTTAACCCCTAGTGTCTCTCACTTAATCTTCTC 1980
 QY 2049 actaagtgaagaatgaagaactgctgtgataagggagagtgaagaaggatatactgtgtagagc 2108
 |||||||
 Db 1981 ACTTAAGTGAAGATGAGAACTGCTGTGATAGGAGAGTGAAGAGGAGATATGTGGTATAGC 2040
 QY 2109 acttgatttaagtgaatgctgtgtgtagctgttccatctctgtgagcgccgttcccta 2168
 |||||||
 Db 2041 ACTTGATTTAAGTGAATGCTGCTGTGATCTTTTCCATTTCTGAGGAGCGCGTTCTCTA 2100
 QY 2169 ataattccaggttgtgtagcgttgaagaagaacttgaatgaaagaagaacttccctctgt 2228
 |||||||
 Db 2101 ATATTTCCAGGTTTGGTAGCGTGGAGAGAACTTTGATGAGAAAGAGAACTTCCCTTCTG 2160

QY 2229 tactgttaacttaaaataatagctcctgatacaagta 2268
 |||||||
 Db 2161 TACTGTTAACTTAAATAATAGCTCTGATTCAAAGTA 2200
 RESULT 3
 AX045253
 LOCUS AX045253 2217 bp DNA linear PAT 24-NOV-2000
 DEFINITION Sequence 5 from Patent WO006727.
 ACCESSION AX045253
 VERSION AX045253.1 GI:11343804
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 2217)
 AUTHORS Adolf, G., Heider, K.H. and Sommergruber, W.
 TITLE Tumour-associated antigen
 JOURNAL Patent: WO 006727-A 5 09-NOV-2000;
 Boehringer Ingelheim International GmbH (DE)
 FEATURES
 source Location/Qualifiers
 1..2217
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 5'UTR 845..1744
 CDS /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC17337.1"
 /db_xref="GI:11343805"
 /translation="MKRLTKLGYCELCMPLEPTVYSMMKKSPFEKAVKAIISCF
 PNVFIASKLVRVYVASWSRVQADLMCEMDLQSSVPMKFLMTGCTDPPKSNAMWQ
 ALKMLNGRSMSEVPEPKRETRMKYHEFVVDLTLNKKKDDPPYNIPTGNAIY
 VASRDFVQVILNPKRSOOLIEWYKDFYSDEHLMAFLRARMPSPVNPYDLSDM
 TSIARLVKIKQGHGEBIDKGAIPAPSGIHQRIICVYGADILNMILQNHILANKRDPK
 VDNALQCLEELRYKAIYTEL"
 1745..2217
 3'UTR
 BASE COUNT 600 a 498 c 547 g 572 t
 ORIGIN
 Query Match 94.1%; Score 2183.2; DB 6; Length 2217;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2196; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
 QY 70 gtacagggaactcccttgacttctgtgacctgccccttaactacagcaatttgtctcg 129
 |||||||
 Db 1 GTCACGGGAATCGCCCTGCTACTTGTGACCTGCCCTTTACTACGACATTTTGTCTGG 60
 QY 130 gaagccctgggaatctgtctaatacctatacactgtagtgctggaaggaaacagataaga 189
 |||||||
 Db 61 GAAGCCCTGGGATTCGCTAATACCTATACCTAGGTGCTGGAAGGAAACAGATGAGCA 120
 QY 190 acatgacctaaagagcttctgtcaatgaagaagcaagcttagccttggcaagaatat 249
 |||||||
 Db 121 ACATGACCTTCAAGAGCTTCTGTCAATGAGAGACCAACTGACCTGGCAAGATAT 180
 QY 250 taaaggagagccggaactgttcccttggacatcttgaatgacagaaataactcttgg 309
 |||||||
 Db 181 TAAAGAGAGCCCTGAAACTTCTCTTGACATCTTATGAATGCACAGAAATACCTTTTGG 240
 QY 310 aggtttagaagatcaggggagacatggttgcacattgtctgcacaggaacacccagtc 369
 |||||||
 Db 241 AGGGTTAGAAAGATCAGGGGACATGTTGTTACATTTGTGCGACGGAACACCGCAGTC 300
 QY 370 ttcaacttggaaacaagaatacagccttgtgaagaagatcattccctaagcaggagaagctca 429
 |||||||
 Db 301 TTCACCTTGGGACACAGATCACGCTTGTGAAGAGATCATCCCTAACAGAGAGAGACTA 360
 QY 430 ctaagaagatgttctcctcccaacttccctgtgcggtgtccacacttctccacttc 489
 |||||||
 Db 361 CTAAAGGATTGTACTCTTCACCTTCCCTGTGGGTGCTCCACCTGTCTCCATTTCT 420

OY	430	gtgcagatggttcaatggaagagactgtccagctgacatctgtggtctctggtctgc	549
Db	421	gtgcagatggttcaatggaagagactgtccagctgacatctgtggtctctggtctgc	480
OY	550	tatatgtctgtgcgcacactgtgctctcgaaacttcttcaagtgtgaagtgtacctgc	609
Db	481	tatatgtctgtgcgcacactgtgctctcgaaacttcttcaagtgtgaagtgtacctgc	540
OY	610	caattggtctggtggtccaaaggaatccaaagccagtaattggaatatcttgaatat	669
Db	541	caattggtctggtggtccaaaggaatctcaaaagccatctgaaggatattcttgatat	600
OY	670	ttctctgaaacttccagaagaaggtctatatactgtgtccaggggttcaaccgagggacca	729
Db	601	ttctctgaaacttccagaagaaggtctatatactgtgtccaggggttcaaccgagggacca	660
OY	730	gagcgaatgcttcaaggtctatctgaaacactggaaggtcagaagaagcgaagccttc	789
Db	661	gagcgaaggtcttcaaggtctatctgaaacactggaaggtcagaagaagcgaagccttc	720
OY	790	acagaaccaccaactcctctccctccacagaagctgtggaacactcaaggtcgaagaag	849
Db	721	acagaaccaccaactcctctccctccacagaagctgtggaacactcaaggtcgaagaag	780
OY	850	ttctacagttccccaactgagcaagaagaagtggaattccctcatgacactctatgtg	909
Db	781	ttctacagttccccaactgagcaagaagaagtggaattccctcatgacactctatgtgtg	840
OY	910	attcaagaagaatltgaaactltgaaaggtactctgcagctgtgtatgccccctcaagac	969
Db	841	attcatgtgaagattgaaacttggaaagcgtactgcagctgtgtatgccccctcaagac	900
OY	970	ataatacgtgtccatgtgagtg-agaagttcccaagaacttccaagaagcgtctaaagc	1028
Db	901	ataatacgtgtccatgtgagtgaaagagtcctccagaagacttccaagaagcgtctaaagc	960
OY	1029	aattattctgtcttcccaatgtcttcatagccagttaagctgtgttcgggtgttatgc	1088
Db	961	aattattctgtgttcccaatgtcttcatagccagttaagctgtgttcgggtgttatgac	1020
OY	1089	ctctctgttccaaggtgtcaagcttgacactcaactgcatgtgagaagactgtctcagaagctcagt	1148
Db	1021	ctctctgttccaaggtgtcaagcttgacactcaactgcatgtgagaagactgtctcagaagctcagt	1080
OY	1149	gcgcgtgaaataacttctctgtatacatgtgtggacgcagcttctctataaagaacatgcaga	1208
Db	1081	gcgcgtgaaataacttctctgtatacatgtgtggacgcagcttctctataaagaacatgcaga	1140
OY	1209	gatgtgtcagaagcttccaagatgttgaatggaggaatagaatgaggttaaggtacccc	1268
Db	1141	gatgtgtcagaagcttccaagatgttgaatggaggaatagaatgaggttaaggtacccc	1200
OY	1269	taagcacaaagaagaacccgcgtgaaatacataactltgagtgagtgtgagagacatatcact	1328
Db	1201	taagcacaaagaagaacccgcgtgaaatacataactltgagtgagtgtgagagacatatcact	1260
OY	1329	aaccacaagaagaaggaatccctccctctataatctaactatgtttaaagggaatgtga	1388
Db	1261	aaccacaagaagaaggaatccctccctctataatctaactatgtttaaagggaatgtga	1320
OY	1389	catgtgtgtctcccgagaattgtgtccaacatgtttagaagaacccaatcccaacact	1448
Db	1321	catgtgtgtctcccgagaattgtgtccaacatgtttagaagaacccaatcccaacact	1380
OY	1449	gatttgaatgtgtaaaagacatatatagccagaatgaacacccctgtggccaaccttcaagc	1508
Db	1381	gatttgaatgtgtaaaagacattatagccagaatgaacacccctgtggccaaccttcaagc	1440
OY	1509	tgcacggtgtgactgtcgtgtgttccccaaccccccaagtaagaacatctcagacatgc	1568
Db	1441	tgcacggtgtgactgtcgtgtgttccccaaccccccaagtaagaacatctcagacatgc	1500

QY	1569	ttctatgtccaggcttggtctcaagtgggcagggtcatgtaggagagacatcoga	1569	aaaggtgtctcc	1628
Db	1501	ttctattgtccaggcttggtctcaagtgggcagggtcatgtaggagagacatcoga	1501	aaaggtgtctcc	1560
QY	1629	ttatgtccctctctctgtgaatccaccacgctatctctgctttatgggtctgggagctt	1629		1688
Db	1561	ttatgtccctctctctctgtgaatccaccacgctatctctgctttatgggtctgggagctt	1561		1620
QY	1689	gaattggagatgcttccaaaccatcacctgtgttgccaaacaagtttgagcccaaggtatgata	1689		1748
Db	1621	gaattggagatgcttccaaaccatcacctgtgttgccaaacaagtttgagcccaaggtatgata	1621		1680
QY	1749	taatgctcttcagtgcttgaagaatatacctaagcttataagccatctatggagcttgaact	1749		1808
Db	1681	taatgctcttcagtgcttgaagaatatacctaagcttataagccatctatggagcttgaact	1681		1740
QY	1809	ttgagacacactatgagagcgtttgctaacctgtggggcaagagacatgtacaacatgtctca	1809		1868
Db	1741	ttgagacacacactatgagagcgtttgctaacctgtggggcaagagacatgtacaacatgtctca	1741		1800
QY	1869	gaacttgctcgggagacagtggtgggtgggagaccagggctttgcaatccgtgtgacatcttag	1869		1928
Db	1801	gaacttgctcgggagacagtggtgggtgggagaccagggctttgcaatccgtgtgacatcttag	1801		1860
QY	1929	gataagaggagcctgcattagatctgtgggtgaagtagatccttggccttgcaaatgtctgcc	1929		1988
Db	1861	gataagaggagcctgcattagatctgtgggtgaagtagatccttggccttgcaaatgtctgcc	1861		1920
QY	1989	tggggtgaatgctgctgtgtctctctccacccttaacccttagtcttcctccactaacttctc	1989		2048
Db	1921	tggggtgaatgctgctgtgtctctctccacccttaacccttagtcttcctccactaacttctc	1921		1980
QY	2049	actaaagtgaagaatgagaactcgtctgtgataaggagagtgaaaggaagatatgtgtgtagagc	2049		2108
Db	1981	actaaagtgaagaatgagaactcgtctgtgataaggagagtgaaaggaagataratgtgtgtagagc	1981		2040
QY	2109	actgatattcagttbaatcctcctgcctgcgtgaagctttccaatctgtgtgagctgccttcccta	2109		2168
Db	2041	actgatattcagttbaatcctcctgcctgcgtgaagctttccaatctgtgtgagctgccttcccta	2041		2100
QY	2169	ataaattccaggttctgttagcgttggagaggaactttgatgtgaaaggaacttccctctctg	2169		2228
Db	2101	ataaattccaggttctgttagcgttggagaggaactttgatgtgaaaggaacttccctctctg	2101		2160
QY	2229	tactgttaacttaaaaaataaatagcttccctgataccaagta	2229	2268	
Db	2161	tactgttaacttaaaaaataaatagcttccctgataccaagta	2161	2200	

RESULT	4
LOCUS	BC017032
DEFINITION	BC017032 Homo sapiens, glucocorticoid (N-acetyl) transferase 3, mucin type, clone MGC:9086 IMAGE:3851937, mRNA, complete cds.
ACCESSION	BC017032
VERSION	BC017032.1 GI:16877561
KEYWORDS	MGC.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 2106)
TITLE	Strausberg,R
JOURNAL	Direct Submission
REMARK	Submitted (05-NOV-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
COMMENT	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc.

QY 1444 caactgataatgggtaaaagacattatagccagatgaacacccctcgtggccaccct 1503
 |||||||
 Db 1302 CAATGATTAATGGGTAAGAAAGACATTATAGCCAGATGAACACTCTGGGCAACCTT 1361
 |||||||
 QY 1504 caagcgtcaagcgttggaatgctgtctgtcccaaccacccaagtgcacatctcagac 1563
 |||||||
 Db 1362 CAGGTCACAGGTCGTAAGCCCTGCTGTGTCCCAACACCCCAAGTACGATCTCGAC 1421
 |||||||
 QY 1564 atgactctatctgcacagcctggtcctcgaatggcaggtcatgaggaagacatcgataaggt 1623
 |||||||
 Db 1422 ATGACTCTTAATGCGCAGGCTGCTCAAGTGGCAGGCTCATAGGAGACATCGATAAGGCT 1481
 |||||||
 QY 1624 gctccctaatgctccctcgtctcttgaaatcccaacagcgtatctgtcgtttatggtcgtgg 1683
 |||||||
 Db 1482 GCTCCTTAATGCTCCCTCTCTGGAAATCCACAGGCGCTATCTGCTTAATGGGGCTGGG 1541
 |||||||
 QY 1684 gacttgatggatgctctcaaaacatcacctgtgtgcacaaagtgttgacccaaggtta 1743
 |||||||
 Db 1542 GACTTGAAATGGATGCTTCAAAACATCCTGTTGGCCAAAGTTGACCCAAAGGTA 1601
 |||||||
 QY 1744 gatgataatgctctcctcagtgcttgaaagaataccctacgttataaagccatctatggact 1803
 |||||||
 Db 1602 GATGATTAATGCTCTTCAAGTCTTGAAGAAATACCTAGTTATAAGGCCATCTATAGGACT 1661
 |||||||
 QY 1804 gaactttgagacacactatgaagcgttgtctacccgtgtgggcaagagcatgtacaaat 1863
 |||||||
 Db 1662 GAACCTTGAGACACACATATGAGACCGTTGCTACTGTGGGGCAAGACATGTACAAACAT 1721
 |||||||
 QY 1864 gctcagaactgtcgtggagacagtggtgggtggagacagggcttgcaattcgtgtacc 1923
 |||||||
 Db 1722 GCTCAGAACTGCTGGGACAGTGTGGGTGGGAGACAGGCGCTTGGCAATTCGTGGCATCC 1781
 |||||||
 QY 1924 tttagaataagagggcctcctatagattgtcgtggaatagatcttgccttgcaaatg 1983
 |||||||
 Db 1782 TTTAGATTAAGAGGGCGTCAATTAATGATGTGGGTAAGTAAATCTTTGCTTGCATAATG 1841
 |||||||
 QY 1984 ctgcctgggtgaatgctgtctgtctcaccacccctaaacccctagtagttcctccactact 2043
 |||||||
 Db 1842 CTGCTCGGGGAATGCGTGTGTCTCTCAACCCCTAAGCTTGAATGTTCCCTCCACTACT 1901
 |||||||
 QY 2044 ttcccaataagtggaatggaatgctgtgtgataaggagagtggaaggagatgtggt 2103
 |||||||
 Db 1902 TTCTCACTAAGTGAATGAGAACTGCTGTGATAGGAGAGTGAAGAGGATATGTGCT 1961
 |||||||
 QY 2104 agagcacttgatcagttgaatgacccctcgtgtgtagctttccatctcgttgagcgt 2163
 |||||||
 Db 1962 AGACACTTGTATTCAGTTGAATGCTGCTGTGATGCTTTTCCATTTCTGTGAGCTGTGCGT 2021
 |||||||
 QY 2164 tctctaataatccaggtttgtgtagcgtgtgagaggaacttgatgagaaagaaacctcc 2223
 |||||||
 Db 2022 TCCATAAATAATCCAGGTTTGTGTAACGTGGAGGAACTTGATGGAAGAAAGAAC -TTCC 2079
 |||||||
 QY 2224 ttctgtactgtaacttaaaataaata 2251
 |||||||
 Db 2080 TTCTGTACTGTTAATAATAATAATA 2107
 |||||||

RESULT 6
 AF102542 2128 bp mRNA linear PRI 31-JAN-1999
 LOCUS AF102542 Homo sapiens beta-1, 6-N-acetylglucosaminyltransferase mRNA,
 DEFINITION complete cds.
 ACCESSION AF102542
 VERSION AF102542.1 GI:4204683
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2128)
 AUTHORS Yeh,J.C., Ong,E. and Fukuda,M.
 TITLE Molecular cloning and expression of a novel beta-1,
 6-N-acetylglucosaminyltransferase that forms core 2, core 4, and I

JOURNAL J. Biol. Chem. 274 (5), 3215-3221 (1999)
 MEDLINE 99115671
 REFERENCE 2 (bases 1 to 2128)
 AUTHORS Yeh,J.C., Ong,E. and Fukuda,M.
 TITLE Direct Submission
 JOURNAL Submitted (28-OCT-1998) Glycobiology Program, The Burnham
 Institute, 10901 N. Torrey Pines Rd., La Jolla, CA 92037, USA
 FEATURES
 source
 1. 2128
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="brain"
 /dev_stage="fetus"
 354. 1670
 /note="glycosyltransferase; core 2 Gnt-M"
 /codon_start=1
 /product="beta-1, 6-N-acetylglucosaminyltransferase"
 /protein_id="AADI0824.1"
 /db_xref="GI:4204684"
 /translation="MYOMRRLCOLHYLMALGCMYLMATVALKISFRLKCDSDHGLIES
 RESOQYCNILYNFLPKPKRSINSGVTRGDOEAVILOAILNNLEVKKKREPETH
 YLSTPRDCHEFRAREKEIOEPISKEVEPIAYSMVTHKERIKETRLRAVAPONTY
 CVHYDEKSPETREKAVKATISCPNPNFASKLVRYVYASMSRQADLNCEMDLOSIV
 PMVYFLNTGCTGTPPIKSNABEVOALKMLGRNMSSEVPPKHEKTKYHFEVVDTL
 HLNRKKDDPPYNLTMFTGNAYIVASRDEVOHLKPKRSOOLIEWVKDYSPDEHLMA
 TLORARMGSGVSNHPRKYDISMTSIALRVKQGHGDDIDKGPVAPCSGIIHQRAICV
 YGADLNWMLQNHLLANKFPRVDNALQCELYRYKATVGTGL"
 BASE COUNT 569 a 477 c 526 g 556 t
 ORIGIN
 Query Match 85.8%; Score 1990.8; DB 9; Length 2128;
 . Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2004; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 244 agatataaagagagcctgaacatgctccttgacacttataatgaatgcaagaataacc 303
 |||||||
 Db 102 AGATATTAAGAGAGAGGAGGCTGAACCTGTTCTTGACATCTTAATGAATGTCAGAAATATCC 161
 |||||||
 QY 304 tttaggggttggaatgaatgaaggagacatgtgtgtacattgtgtgcacggaaacccg 363
 |||||||
 Db 162 TTTTGGAGGGTTGAAGAAATAGGGAGACATGTTGTTACATTTGCTGCGCACGGAACACCG 221
 |||||||
 QY 364 ccagcttcaccttggaaacgaatcaacgaccttgtgaagaagatataccctaagcagaagag 423
 |||||||
 Db 222 CAGGTCTTCACTTGGAAAGAGAAATACAGCCTGTGGAAGAATATCATCCTTAAGCAGAGAG 281
 |||||||
 QY 424 aagctactaaagatgtgtctcctccctccacccctcctgtgtcgtgtccactgtctcc 483
 |||||||
 Db 282 AAGCTACTAAAGGATGTGTCTCTCCACCTTCCCTGCTGCTGCTCCACCTGTCTCC 341
 |||||||
 QY 484 cattctgtgacgagtgtgttaactggaagaagactgtgcacagctgattactgtgtgacctg 543
 |||||||
 Db 342 CATTCTGTGACGATGTTCAATGGAAGAGACTGCGCAGCTGCACTTAATCTGTGTGGGCTGTG 401
 |||||||
 QY 544 ggcctctataatgctcgtcgtccacgtgtgtccttgaacattccttcaggttgaaagtgtgac 603
 |||||||
 Db 402 GCGCTGTAATGCTGCTGGCGCACTGTGCTGTAATCTTCTTCAAGTGTGAAGTGTGAC 461
 |||||||
 QY 604 tctgacacactgtgtcgtgagctccagggaaatctcgaagcagactgtgaaatattctg 663
 |||||||
 Db 462 TCTGACACACTTGGGTGTGAGTCCAGGAGATTCMAAGCCAGTACTGTAGATAATATCTTG 521
 |||||||
 QY 664 tataatttccctgaacttccagcaagaaggtctataacactgttcaaggggtccaccgaagg 723
 |||||||
 Db 522 TATATATTCCTGAAACTTCAGCAAGAGGTCTATCAACTGTTCAGGGGTCCACCGAGG 581
 |||||||
 QY 724 gaccacaagcagctgtcctcagctattctgaactccttggaaggttcaagaagaagcagagag 783
 |||||||
 Db 582 GACCACAGAGGAGCTTCAAGGCTATTCTGAAATTAACCTGAGGCTCAAGAGAAAGCCAGAG 641
 |||||||
 QY 784 ccttcacagacaccactactctcctccacagagactgtgagcacttcaaggtgaa 843

```

|||||
Db 642 CTTTACAGACACCCTACTCTCCCTCACAGAGACTGTGAGCATTTCAAGGCTGAA 701
|||
QY 844 aagaagttacatagttccctccatgcaagaagagtgtagttccctatgtcatctc 903
|||
Db 702 AGGAAGTTATACAGTCTCCACTGAGCAAGAAGAGTGGACTCCCTTTCCATACCTCT 761
|||
QY 904 atggtatcatcagagaagatgtaaaacttgaagaagctactcgagctgtatgtccct 963
|||
Db 762 ATGGTATTCATGAGAGAGTTGAAAACCTTGAAAGGCTACTCGAGCTGTATGTCCCT 821
|||
QY 964 cagaacatactgtgtccatgtgtgagtagaagttcccgaaactttcaagaagcgtc 1023
|||
Db 822 CAGAATATATACGTGTCCATGTGATGAGAAAGTCCCAAGAACTTTCAAGAGCGGCTC 881
|||
QY 1024 aagaacattatcttctgtctcccaaatgtctcatatgacgaagtaagttgttggtgtc 1083
|||
Db 882 AAGGCAATATATTTCTGTGCTCCAAATGTCTTCATAGCCAGTAAGCTGGTGGGTGT 941
|||
QY 1084 tatgtctctgtgtccagagtgcaagctgaacctgaactgcatggaagactgtctcagagc 1143
|||
Db 942 TATGCTCTCTGTGTCCAGGTTGCAAGCTGACCTCACTGATGAGAGACTTGTCTCCAGAGC 1001
|||
QY 1144 tcaatgctgtggaabaactctctgaatatagttggaagcgtacttccataaagaagcaat 1203
|||
Db 1002 TCAGTGCCTGTGAAATACCTCTGAATACATGTGGAGCGAGCTTCTTAAAGACCAAT 1061
|||
QY 1204 ggaagatgtgtcagagcttcaagaatgttgaaatggaagaaatagactgtgagcagagta 1263
|||
Db 1062 GAGAGATATGTCCAGGCTCTCAAGATGTGAATGGAGGAATAGCATGTGAGTCAGAGTA 1121
|||
QY 1264 cctcctaagcacaaagaaacccgcttgaaatatacacttgtagttagtgagagacatata 1323
|||
Db 1122 CCTCCTAAGCACAAAGAAACCCGCTGGAATATCATTTGAGGTAGTGAAGACACATTA 1181
|||
QY 1324 caactaacacaaagaagaagatccctccctataactaactatgtttcaagggaaat 1383
|||
Db 1182 CACCTAACCAACAGAGAGAGATCTCCCTTAAATTAATTAATTAATTAATTAATTAAT 1241
|||
QY 1384 ggtatcatgtgtgtcccgagatgtgtcacaacatgtttgaaagaccttaaccata 1443
|||
Db 1242 GCGTACATGTGTGCTTCCCGAATTTTCGTCACACATGTTTGAAGAACCTTAATCCCA 1301
|||
QY 1444 caactgatgtaaatggttaaaagacacttataagccagatgaacaacctctcggccacct 1503
|||
Db 1302 CAACGTATGATGAGTAAAGACACTTATAGCCAGATGAACACCTCTGGGCGACACCTT 1361
|||
QY 1504 caggtgtcagagtgatgtcgtgtcgtctgtcccaaccaccccaagtagacatctagac 1563
|||
Db 1362 CAGCGTGACAGGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1421
|||
QY 1564 atgactctatgtccagagtggtgtcaagtgtcagagtgatgagagagacatcgtaagagtt 1623
|||
Db 1422 ATGACTTCTATTGCTCCAGGCTGTGCTCAAGTGTGCAAGGATCAAGAGAGACATCGTAAGGGT 1481
|||
QY 1624 gctcctatgtctcctgtcgtctgtgaatccacagcgggctatctcgtttatgtggtgtg 1683
|||
Db 1482 GCTCCTATTAGCTCCCTGCTGTGGAATCCACAGCGGGCTATCTGCGTTATAGGGGTGGG 1541
|||
QY 1684 gacttgaattgtatgtcttaaaacatcaactgtgttgccaacaagtttgaaccaaaagta 1743
|||
Db 1542 GACTTGAAATGTGATGCTTCAAAACATCACTGCTGTGCGCAACAAAGTTTGAACCAAGTA 1601
|||
QY 1744 gatgataatgtcttcaagtgcttagaagaataatcctcgttataagggcactatgtgact 1803
|||
Db 1602 GATGATAATGTCTTCTAGTGTAGGAAGATATCATACGTTATAGGCGCATATATGGGACT 1661
|||
QY 1804 gaacttggagaacacatagagagcgttgtactcgttggtgcaagagacatgtatacaacat 1863
|||
Db 1662 GAACTTTGAGACACACTATGAGAGCGTTCTTACTTGTGGGCAAGAGCATGTATCAACAT 1721
|||
QY 1864 gctcagaacttgtggaacagtggtgtggtggaagaccagggcttgcgaatctgtgacatc 1923
|||

```

```

Db 1722 GCTCAAGACTTGTGAGAGAGTGTGGTGAGAGACCAGGCTTTGCCATTCGTGCATCC 1781
|||
QY 1924 tttagataagagggtgtgtatagatgtgtgtagtagtagatcttctgttgcaaatg 1983
|||
Db 1782 TTTAGATATAGAGGGGTGTGTATTAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1841
|||
QY 1984 ctgacctgtgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2043
|||
Db 1842 CTGCGGTGTGTAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1901
|||
QY 2044 ttctactaagtgtgagaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2103
|||
Db 1902 TTTCTCACTAAGTGTGAGTAATGAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1961
|||
QY 2104 agagcactgtatctcaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2163
|||
Db 1962 AGAGCACTGTGATTTCAATGTAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2021
|||
QY 2164 tccataaatccaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2223
|||
Db 2022 TCTTATATATTCAGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2079
|||
QY 2224 ttctgtactgttcaacttaaaataaata 2251
|||
Db 2080 TTCTGTACTGTTAACCTTAATAATAATA 2107
|||

```

RESULT 7

AC092755 187275 bp DNA linear HTG 19-JAN-2002

LOCUS Homo sapiens chromosome 15 clone RP11-361D15 map 15q22, ***

DEFINITION SEQUENCING IN PROGRESS ***, 2 ordered pieces.

ACCESSION AC092755 AC022480

VERSION AC092755.3 GI:18249989

KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEPIN.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 187275)

AUTHORS Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 187275)

Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R., Pate,D. and Hood,L.

Sequencing of human chromosome 15 D15S146-D15S117 region

Unpublished

2 (bases 1 to 187275)

Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R., Pate,D. and Hood,L.

Direct Submission

Submitted (26-JUN-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA

On Jan 19, 2002 this sequence version replaced gi:15022677.

----- Genome Center

Center: Multimegabase Sequencing Center

Web site: http://chroma.mbt.washington.edu/msg.mw

Contact: leerowen@systemsbiology.org

----- Summary Statistics

Sequencing vector: pUC18; 108752

Chemistry: Dye-terminator Big Dye; 90% of reads

Chemistry: Dye-primer Big Dye; 10% of reads

Assembly program: Phrap; version 0.990399

Insert size: -----; agarose-tp

Quality coverage: ---x in Q20 bases; sum-of-contigs

----- Sequence Quality Assessment: -----

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 32680: contig of 32680 bp in length
* 32681 32780: gap of unknown length
* 32781 187275: contig of 154495 bp in length.
Location/Qualifiers
1. 187275

FEATURES
Source
1. 187275
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q22"
/clone="RP11-361D15"
/clone_1db="RPCT human BAC library 11"
/note="This clone overlaps RP11-112N19 and RP11-568G20"

BASE COUNT 53688 a 38780 c 38083 g 56624 t 100 others
ORIGIN

Query Match 81.28; Score 1883.8; DB 2; Length 187275;
Best Local Similarity 99.96; Pred. No. 0;
Matches 1885; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 433 aagatgtgtcctccaccctccctcctgtgtcgtctccaccgtgtccattctgtg 492
Db 40055 AAGATGTGTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 40114
Qy 493 aagatgtgtcctccaccctccctcctcctcctcctcctcctcctcctcct 552
Db 40115 ACGATGTGTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 40174
Qy 553 atgtgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 612
Db 40175 ATGT 40234
Qy 613 ttgt 672
Db 40235 TTGGGT 40294
Qy 673 ctgaactctcagaagaagagctctatcaactgttcagggtgtcaccgagggaccagag 732
Db 40295 CTGAACCTTCAGCAAGAGAGGTCTATCAACTGTTCAGGGGTCAACCGAGGACCAAGAG 40354
Qy 733 gcaatgtctgaagctatcttgaataactgtgaggtcaagaagaagcgagcccttca 792
Db 40355 GCAATGTCTCAGGTATTTCTGAATPACCTGAGGTCAAGAGAGCGAGACCTTTTCA 40414
Qy 793 gacacccactaccctccctcaccagagactgttagcactcagaaggtcgaagaagtc 852
Db 40415 GACACCCACTACTCTCTCCCTACAGAGACTGTGAGCACTTCAAGGCGTGAAGAGATTC 40474
Qy 853 atacagttcccaactgtgacaagaaggttgagttccctattgatactatactgtgtg 912
Db 40475 ATACAGTTCCCACTCAGAGCAAGAGAGGTGAGTCCCTATTTGCACTCTATGTGAT 40534
Qy 913 caltaagaagattgaagaacttgaagaagctactgagctgtgtgtgtgtgtgtgtgt 972
Db 40535 CATGAAGAAGATTGAAGACTTGAAGAGGTACTGAGGTGTGTGTGTGTGTGTGTGT 40594
Qy 973 tactgttccatgt 1032
Db 40595 TACTGTGTCCATGT 40654

Qy 1033 attcttgcctcccaatgtctcctcctcctcctcctcctcctcctcctcctcctc 1092
Db 40655 ATTCTTGTCTTCCCAATGTCTTCTTATGACCGATGAGTGTGTGTGTGTGTGTGTGT 40714
Qy 1093 tgtgtccaggtgtcgaagctgtgacctcctcctcctcctcctcctcctcctcctc 1152
Db 40715 TGTGTCCAGGGTGTCAAGTGTGACCTCACTCACTCACTCACTCACTCACTCA 40774
Qy 1153 tggaaactcctcgaataactgtgagagagagagagagagagagagagagagagag 1212
Db 40775 TGGAAATACCTTCCGATATACATGTGGAGACGACTTCTTAAAGACATGCAAGATG 40834
Qy 1213 gtccagctcctcgaatgttgaatgtggaaggaatgaatgtgagtcagtcctcctc 1272
Db 40835 GTCCAGCTCTCAAGATGTGGAATGGAGGAATGTGATGATGATGATGATGATGATG 40894
Qy 1273 cacaagaagaacccgtgtgaatatcactttgaggtatgtgagagagagagagagag 1332
Db 40895 CACAAGAAGAACCCGTGTGAATATACATTTGAGGTATGTGAGAGACATTAACACTTAAC 40954
Qy 1333 aacaagaagaagatcctcctccctcctcctcctcctcctcctcctcctcctcctc 1392
Db 40955 AACAGAAGAAGATCCTCCCTCTTATATTTAATTTAATTTAATTTAATTTAATTT 41014
Qy 1393 gtgtgtcccgagatcctcctcctcctcctcctcctcctcctcctcctcctcctc 1452
Db 41015 GTGTGTCCCGAGATTCGTCTCCCAACATGTGTAAGAACTTAAATCCCAACACTGAT 41074
Qy 1453 gaatggtaaaagacactttaagccagatgaacactcctcctcctcctcctcctcctc 1512
Db 41075 GAATGGTAAAGACACTTATAGCCAGATGAACCTCTGAGGACCTTACACCTGCA 41134
Qy 1513 cgt 1572
Db 41135 CGT 41194
Qy 1573 attgccaggtgtgtcgaatgtgcaaggtgtcgaaggtgtcgaaggtgtgtgtgtgt 1632
Db 41195 ATTGCCAGGCTGTGCAAGTGTGCAAGGTGTGCAAGGTGTGCAAGGTGTGCAAG 41254
Qy 1633 gctccctgt 1692
Db 41255 GCTCCCTGT 41314
Qy 1693 ttgtgtcctcacaac 1752
Db 41315 TGT 41374
Qy 1753 gctctcagtgcttagaagaatcctcagtgatataagggcatctatgtgagctgaacttga 1812
Db 41375 GCTCTCAGTGTCTTAGAAGATACCTGATTAAGGCCATCTATGTGGACTGAACCTTGA 41434
Qy 1813 gacacactatgagagagcttctcactcgttgggcaagaagagatgaacaactgtccagaac 1872
Db 41435 GACACACTATGAGAGCCTTGTCTACTGTGTGGGCAAGACATGTACAAACATGCTCAAC 41494
Qy 1873 ttgtgtggaacagtggt 1932
Db 41495 TTGTGTGGAGACGT 41554
Qy 1933 agagagctgtcctatagattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1992
Db 41555 AGAGAGCTGTCTATTAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 41614
Qy 1993 tgaatgtcgt 2052
Db 41615 TGAATGTCTGT 41674
Qy 2053 agtgaagaatgaac 2112
Db 41675 AGTGAAGATGAGACTGT 41734
Qy 2113 gattcagttgaatgtcgt 2172


```

RESULT      9
AF038650
LOCUS       AF038650                1317 bp    mRNA    linear    PRI 25-MAR-1999
DEFINITION  Homo sapiens core 2/core 4 beta-1,6-N-acetylglucosaminyltransferase
            (C2/4Gnt) mRNA, complete cds.
ACCESSION   AF038650
VERSION     AF038650.1
KEYWORDS    AF038650.1  GI:4511880
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1317)
AUTHORS     Schwenek,T., Nomoto,M., Lavery,S.B., Merkh,G., van Kessel,A.G.,
            Bennett,E.P., Hollingsworth,M.A. and Clausen,H.
            Control of O-glycan branch formation. Molecular cloning of human
            cDNA encoding a novel beta1,6-N-acetylglucosaminyltransferase
            forming core 2 and core 4
            J. Biol. Chem. 274 (8), 4504-4512 (1999)
            99143102
JOURNAL     MEDLINE
REFERENCE   2 (bases 1 to 1317)
AUTHORS     Schwenek,T. and Clausen,H.
            Direct Submission
            Submitted (28-NOV-1997) Dental School, Faculty of Health Sciences,
            Copenhagen University, Noerre Alle 20, Copenhagen 2200, Denmark
FEATURES
    source
        1..1317
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            1..1317
            /gene="C2/4Gnt"
            1..1317
            /gene="C2/4Gnt"
            /note="core 2/4-Gnt"
            /codon_start=1
            /product="core 2/core 4
            beta-1,6-N-acetylglucosaminyltransferase"
            /protein_id="AAD21525.1"
            /db_xref="GI:4511881"
            /translation="MYQWRKLCQHYMALGCMYLAATVLAITLSPRLKQSDHLGIES
            RESQSOYCRNIIYNPLAKPRKRISNGSVTRGQEOVLQALIKLNLEVKKKREFPTDTH
            YLSLRKDCHEFKERKFIQPLSKSEVEFPIATSMVTHEITENFERLIRVAVAPQNIY
            CVHVDKSPETFEKAVKAIISCPNVEIASLRVYVSVSRYQADLNCMEDLQSSV
            PMKYPLNTGCTDPEIKSNMEVQALKMLNRMSESEVPERKHETRMKYFEVVRDL
            HLTNRKKDPPRYMLTMTGNAYIVASDFQVHLKPKRSQDLIEWKDTYSPDEHMA
            TLQRRMMEGVSVPNHPKDYDSDMTSIALRVKQGHGSDIDKGA.PVACSGIHQRAICV
            YGAGDINMMLQNHHLANKFDPKVDNALQCLEEYLRKRYKTYGTEL"
BASE COUNT  359 a      307 c      322 g      329 t
ORIGIN
Query Match      56.8%; Score 1317; DB 9; Length 1317;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 736 gtcctcagcgtatctctgaataaccttgaggctcaagaagcgcagcccttcacagac 795
    |||||
DB 241 GTGCTTCAGGCTATTTTGAAATTAACCTCGAGCTCAAGAAGAAGGAGCGCTTTCACAGAC 300
QY 796 acccaactacctccctccaccagagactgtgagcaactcaaggctgaaggaagttcata 855
    |||||
DB 301 ACCCAGTACCTCTCCCTCACAGAGACTGTGACACTTCAAGCGTGAAGAAGATTCATA 360
QY 856 cagtcccaactgagcaagaagaggtgagttccctatgcatactatggtgattcat 915
    |||||
DB 361 CAGTTCCTACTGACCAAGAAGAGAGTGAGATTCCCTATTGCATATCTATGATTCAT 420
QY 916 gagaagaattgaacttgaaaggctactgcgagctgtgtatgccccctcagaacataac 975
    |||||
DB 421 GAGAGAGATTGAAAACTTTGAAAGGCTACTGCGAGCTGTATGCCCCCTCAGAACATATAC 480
QY 976 tggctcatgtgtagtgaagagtcctcccaagaacttcaagaagcggtcgaagaattatt 1035
    |||||
DB 481 TGTGTCATGTGTGATGAGAGTCCCCAGAAACTTCAAAAGAGCGGTCAAAAGCAATTATT 540
QY 1036 tcttgctcccaatgctctcatagccagtaagctgtgtcggtgttattgctcctcgg 1095
    |||||
DB 541 TCTTGCTTCCCAATGCTCTCATAGCCAGTAACTGGTTGCGGTGTTATGCTCCTCG 600
QY 1096 tccaaggctgaagctgagccctcaactgcagaggaactgtgccaagctcagtgccgtg 1155
    |||||
DB 601 TCCAGGCTGCAAGCTGACCTCAACTGCATGAGACTGCTCCAGACTCAGTCCCTCG 660
QY 1156 aaactactcctgaatacatgttgagagagacttccctctaaaggcaatgtagatgtgc 1215
    |||||
DB 661 AAATACTTCCTGATTAATGTGGACGAGCTTTCCTTAAGAAGCAATGCAAGATGTGC 720
QY 1216 caagctctcaagatgttgaaatgaggaagatagacatgagctcaagagtaacctcctaagcac 1275
    |||||
DB 721 CAGGCTCTCAAGATGTTGAATGGAGAGATAGCATGTAGTCAGAGTACTCTCAAGCAC 780
QY 1276 aaagaacccgcgtggaataatcactttgaggtgtagtagagacaatacctaaccacac 1335
    |||||
DB 781 AAAAAGAAACCCGCTGGAATATCACTTGTGAGTGTGAGAACACATTAACCTCAACCAAC 840
QY 1336 aagaagaagatcctccccccttaactatgactatgcttcaaggaatgtgcatatgtg 1395
    |||||
DB 841 AAGAGAAGAGATCTCCCTTATTAATTTAATTAATTTACAGGGAATGCTATTCATGTG 900
QY 1396 gcttcccgagatctgctcacaacatgcttgaagaaccctaaatcccaacaactgattgaa 1455
    |||||
DB 901 GCTTCCCGAGATTTCTGTCACAATGTTTGAAGAACCCTTAATCCCAACAATGATTGAA 960
QY 1456 tgggttaaaagacacttaagcccaagatgaacacctctgggccaacctcaagctgacgg 1515
    |||||
DB 961 TGGGTAAAGACACTTATTAACCCAGATGAACACCTCTGGGCCACCTTCAGCGTGCACGG 1020
QY 1516 tggatgctcgtgctctgtctccaaccaccaccaagtaagacatcacaagacttatt 1575
    |||||
DB 1021 TGGATGCTCTGGCTCTGTTCCCAACACACCCTCAAGTAGACATCTCAGCATGACTTTCATT 1080
QY 1576 gccagagcttgatcgaatgagcaggtgcatgaggagagacatcgataaagggtgctctatgct 1635
    |||||
DB 1081 GCCAGGCTGGTCAAGTGGCAGGGGTCAATGAGGAGACATCATTAAGGGTGCCTTATGCT 1140
QY 1636 ccctgctctggaatccaacagcggtactatcgcttatabgggctggggaactggaattgg 1695
    |||||
DB 1141 CCGTGTCTGGAATCCACACGCGGGCTATCTGCTTATGGGCTGGGGACTTGAATTTGG 1200
QY 1696 atgcttcaaaaacataccctggttgccaacaagtgtgaccacaagtgagatgataagct 1755
    |||||
DB 1201 ATGCTTAAACCATCACTCCGTTGGCCCAACAAGTTTGAACCAAGAGTAGATGAATAGCT 1260
QY 1756 cttaagtgttaagaagatacctaagcttaaggaagccatgtaggaactggaacttga 1812
    |||||
DB 1261 CTTCAAGTCTTAGAAGAACTACTAATGAGGCCATCTATGGGACTGAACATTGGA 1317

```

```

RESULT 10
AF318573      108873 bp   DNA       linear   VRL 18-MAY-2001
LOCUS        Bovine herpesvirus 4 long unique region, complete sequence.
DEFINITION   AF318573 AF212121
ACCESSION    AF318573.1 GI:12802528
VERSION      AF318573.1 GI:12802528
KEYWORDS
SOURCE       Bovine herpesvirus 4.
ORGANISM     Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
              Gammaherpesvirinae.
REFERENCE    1 (bases 1 to 108873)
AUTHORS      Zimmermann,W., Broll,H., Ehlers,B., Buhk,H.-J., Rosenthal,A. and
              Goltz,M.
TITLE        Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and
              Identification of an Origin of DNA Replication
JOURNAL      J. Virol. 75 (3), 1186-1194 (2001)
PUBMED       11152491
REFERENCE    2 (bases 1 to 108873)
AUTHORS      Zimmermann,W., Broll,H., Ehlers,B., Buhk,H.-J., Rosenthal,A. and
              Goltz,M.
TITLE        Direct Submission
JOURNAL      Submitted (03-NOV-2000) Project P24, Robert Koch-Institut, Nordufer
              20, Berlin D-10555, Germany
FEATURES
source       1..108873
              /organism="Bovine herpesvirus 4"
              /db_xref="taxon:10385"
              /note="Long unique region; LUR"
              complement(445..633)
              /note="ORF Bo1"
              /codon_start=1
              /evidence=not_experimental
              /product="hypothetical protein pbol"
              /protein_id="AAK07921.1"
              /db_xref="GI:12802529"
              /translation="MEGDVSMVAFGRKRTRODLNNSAGNEVCCAVLRKNHNVSAGC
              CAVKRNNSFHHSFLGMPW"
              614..4480
              /note="tegument protein; ORF3 BORP1"
              /codon_start=1
              /product="v-FGAM-synthase"
              /protein_id="AAK07922.1"
              /db_xref="GI:12802530"
              /translation="MDTSPSTQCFEVSRRAPLSMEGOVTOQFTSDQGLPTSLVPLQY
              EOYPMVYNGEBOAVRELVPLDLKFSHGELTFPHSSILKTSRSLTYGVSPSA
              PETALSRLDVAIIQSGSLAPMSMLTQILKIVSILISIKCNLMPSPKIHSTLRAL
              CPDLSLRPRYKGEIVHGAFHTGAPGVLSPPKGDLAGVGFNLNAGSGASGFMT
              POHPQRTIVNQNFETAVIVPDVWIEISITVTLSSGHMTDGLATYTGQLOVAPRD
              LKFLYNSPNQMFNVFSSQLGIASIVSITSPRSSSELPYTRDIDMAMIMRCC
              SLPTGCGFEPPIVSHLPLDSVPEGPICAFGAVCFESRITTDLPETRRKNITVQLGT
              FTSHEDSGREVRCLGALKYLSFLRPOVKGISAVKVCETMAKLLSVCOGTGIGIK
              IYOSALPPLNTGLLPPTASNRQNMNMKQHLNIVAVIVLVAIEATPDMTTIVLK
              AGCAIHCCQYKVLKRTQPHIHLIKDKGEQNVREKISLKRSPSYPLSTINKLOKS
              ENMDDSEMEDTLPSTLYTLEILKHVAGCKDFIVKHIDRLSGRAVOOQIGARDI
              PISDYSILVDSLAAAPDGSBNMATAESLDLPLVNNPVPVPGICSAIGECPES
              TYPRIKGAHVAIVESLNIIVASAPRKRDIDVTCTNITMPOTDSHFGILBELMRGRER
              CSQIGICVPTSCITTSNRGKCLNLSLVRLVITAPCAADVQGLTLPDKEDDSI
              VWLPSTIEHTFPGVISOQLFRDPSGHIINDPLVKKLITVTKHLINHSIISCHVY
              GCGGLITACEMAVAGASIALIVPODEDVPLTSETPGILMVEPRVKSTVOXHE
              TSDIIEYDGRTELVASNTFTVSKERITFRESLNEMVENMRHSTKEOMRTYCEY
              OGCPEDPDKDLOHLTROPSCPHGPKYKHQVANYLLPGTNPDSLLVLAEEAFRN
              LNSTFPTKYIKIYDITNVFGICLIGATINEDATLGDKAISMTTKHNSVVLGELKLI
              NSPDVSLAIGHNCOULEKKEKWIYKPKSPSTWCKENSGMESRWLNFEPENH
              AVALOSMSGLPGLMIOGTHLGFAMHSETYEMWSTGMAVATQYCADISAGPALTP
              QNPATGYTISGLCSADGRHLALHDPGLSNLMQWPHI PKMTPLKYSVPMKRMFLDIH
              IMANKVREMDQPPPHHPDPLNLIKVM"
              4608..4811
              /note="ORF Bo2; BORFA2"
              /codon_start=1
              /evidence=not_experimental
              /product="hypothetical protein pboc2"
CDS

```

```

              /protein_id="AAK07923.1"
              /db_xref="GI:12802531"
              /translation="MASPTQYCHRWVRHGPGRKSAKISYGYKRDYTHQLSKNSAYS
              GMSPOQPCSCADLNKLFVSQCPIS"
              5286..5513
              /note="ORF Bo3"
              /codon_start=1
              /evidence=not_experimental
              /product="hypothetical protein pboc3"
              /protein_id="AAK07924.1"
              /db_xref="GI:12802532"
              /translation="MGFNPVMLYFHPNCNCTPTSLKHGLYIKYISYLIRATFLVRNV
              HIGPPKRDTPRPGGPPSILVQRLAKK"
              5943..9347
              /note="ORF 6"
              /codon_start=1
              /product="single-stranded DNA-binding protein MDBP"
              /protein_id="AAK07925.1"
              /db_xref="GI:12802533"
              /translation="MAKRTAGGOOYEDNIGSKADIPGCGYIYYKKRETPYKEASILG
              NGPFGTSVMSYPLFGLTVEBGPPIYKAVYKAKIDTIVSYKATSYHREVTMEFNIDC
              FTFJPHGGLPOLQETRDLEFTHISPPSNTITNNSLQDLGIDNCLIDENITLGVIT
              EGFERLYHGHMIBPLGVHEEVOISPOHAARKIDLYDEDLFSKLAAPRMKFRHSDVS
              QYLFNSLYTATAQAIYVKDVAIVIQAMESQFVRQHKPKVYHKRKPSSSRGPDGL
              SLIMDSVASELAVSYGLSFIDAPODMSALDYSWPIFYVDCITEEDRLQALHAMNLK
              OSIHVNTQLFSTNMLYLRIOKONPKAKDMLVYNSYLYQHLQHSYSLATODYGO
              PVFGVOSNLSSGTYTTHHLAYASACNPRLAKYCYLQFCQHKSTQSONSYNSY
              VGTASSDMCMVCOGKRPACVINTLFLRLRPRPPTITSHRDPYITGSVGMNDLD
              ILGNFASREKEDGAEQVQKAYTYWOLITJLEKLEGMGIKDTPPEKLVSPISF
              IKVRFDAHADVDREYLFKINCMKNINIIFRETIKSIHHVIOCCNVEOPKCPVFLQ
              IYKSLITFIQDVCLPNCMAYEODNNSMGCPTEWLKMHYOTLWTFNKGCCPDGVLTG
              AELKVMHSDPRCDFEDVAALNGIESPQYVRSRAVLVYPKTIKIKRNIIFESNSG
              SEATQSGFLKSGSKKDNITYGTPIYMKELNTHKIMFPOTKASSLPMTHTSTKQIPL
              ITGVKQDLVALNAVIEYNSKLHSEIDVLDITIPNLITYAKIKLNALLRACGTOFY
              STYLKCVLPVIRVAVGEYPVHLSENEELGEOYLKVQNTAQIVATLKEDEVAOMG
              KLRHFIYPMVINKYTYGNGNNGIFHGCENLGYFRGVDRLMIFEMAFKQSTNAYM
              KRRHFEMPIVDNLIKRTTVPSSTFEVENLRIMLLEBDKDNONIEPDVLELYG
              LGQACADITADDLQIYGEYGIYMSDEILSRQSIISDGVPCESVSFLDEIVYSEE
              QLEPIGLGEORTTSAPTSDPFGASGLSTIAAGKKRKLNCMLSDPL"
              9352..11430
              /note="ORF 7"
              /codon_start=1
              /product="transpore protein"
              /protein_id="AAK07926.1"
              /db_xref="GI:12802534"
              /translation="MARLEAVYQVDPDLSKTSILTFCDPRHIDLGAFQNNKRLKL
              LIDLALVLEKQNTRESCSLSELEHLITINYITMOKLETLNENITKDFPAMHLIS
              DLCCFHKVNLCEFGDCVPIPLSITINDIEFRLNSVFCISRNNALEGELETNF
              LGTFRNTEPIPLDPLYPLSICMCLNTESLIPQGENVFTIMNSINCTHATVAPAE
              PVQGLFEHELHMLNPVGTQNKVQLETRIGNASIEALIDONHIFEKYQHLIEIS
              NLIIYNSGSHSDNPKLSCSQMATLLIQUETBMAHROKVAASHLALTPPIHFPDSYRA
              YPISLRCGGIFITYIMDNIALKQDCSTFLAKSNYKTIHRKNELYRLNNLAPSH
              TKESDLPNTSISQARHTTIGETSEDSIDAARKNNYVLOKVTBKLITDQLETF
              OQKIMCNTLTLRTGMVYVKEAAILKHNFMIRKRFELAPQWHNYVATINELYESKFI
              KNSLYVQLISOEHDSITLOEYSITLGGVGPDEFSPPIELANCLADAKARVPHNK
              MLISEVIMSWEPKQMDIAENAFNYITGKLVNSOOCAMRYITELANCSVILYVRWE
              KDNIVSEPTSPHNLSPKRTFESGIITYTEKAPLFIHDTNGWIFKDLIYALLYHH
              LQLSHKNVL"
              11420..14044
              /gene="gb"
              11420..14044
              /gene="gb"
              /note="ORF 8"
              /codon_start=1
              /product="glycoprotein B"
              /protein_id="AAK07927.1"
              /db_xref="GI:12802535"
              /translation="MYKTIIFALIKVCSNPTTSTTSPSSSTSTSTSTSK
              PSNPTSNSSSLAASPONTSTSKPSTNOGSTPTPIPVTDPLAKNPKYKRVCSASS
              SGELFRDLDDOTCDPKOKKHVEGILVLKKNIPYITFKYRKIRIKATISVYTHGMSO
              AAVTNRDIDISALPYNEISMDRTYHCFSAVAIVINILNTYIDRDSKNSVPLQPAV
              GLTENIRYFSOPLIAVAPGFIYRVRTNENLVVDNMAKRSVEPYTHFTALGDPI
              EISPCNNNSCCTGNSTSRDATKVIENHQVYDVEBRGHPTKDKRILFLKDEEYTS

```


Db 1135 CTCTCGAGCCCTTGTCTGCAACATGCTTAGACAAACCTTAATCCCAATACCTGTTGATG 1194
 Qy 1458 ggttaaagacactatgaccagatgacacaccttggccacccttaagctgaagctg 1517
 Db 1195 GGTAAAGACACATATAGCCCGACGACACCTCTGGGCCACCTTACAGCTGCTCCGTG 1254
 Qy 1518 gatgctgtgctctgttcccaacaccccaagtlacgacatctcagacatgacttattgc 1577
 Db 1255 GATGCCGTGGCTGTCTGCTGACCCCAAAAGTATCATCTCAGACATGACGCTCATCGC 1314
 Qy 1578 caagctgtgtaagtggaaggtcatgagagacatgataaagtgctccttattgctcc 1637
 Db 1315 CAGGCTGTCAAGTGGAGTACACAGAGGAGATGTACAGCTGGGGCGCTTATGACAC 1374
 Qy 1638 ctgtctcgaatccaccagcagcgtatctcgtttatcgggctgggagacttgatg 1697
 Db 1375 CTGCTCTGGAATCATGTGGAGGGCATCTGCATTACGGGGCCGGGACCTGTACTGGAT 1434
 Qy 1698 gcttcaaacacatcactgttggccaaagtttgaaccaaaagtlagatgataagctct 1757
 Db 1435 TCCTCAAGAACATCACCCTTGTGCAAAAGTTGACCCGAGGGGTGATGATACGTCCT 1494
 Qy 1758 tcaatgttgaagaataactaagcttataagggcactctatgggactgaacttgaagac 1817
 Db 1495 GCACTGCTTGAAGAGTACTACATCATAGGCCATCTATGGGACTGAACCTTTGA---- 1549
 Qy 1818 actatgagagcgttctgactcctgtgggcaagagcatgtacaacatgtcagaacttct 1877
 Db 1550 -----GCTGCACGCTGGGCGCAAGGACTGTGCCAACAATGCCAGGGCGTACT 1555
 Qy 1878 gggacagatgt-ggglyggagaccagggcttgcacatcgttgcacatccttgaagaag 1936
 Db 1596 GGTACTGTGTGGGTGGGAGCGAGGGCTGTGGAATCATGATGCCCGCAGGGTCAAGAG 1655
 Qy 1937 ggcctgcatatgattgtgggtaagtag 1963
 Db 1656 GGTGCTTCATGGGTGGGTGACGATGTGG 1682

RESULT 12
BHVAREGP 4179 bp DNA linear VRL 06-JUL-1995
LOCUS Bovine herpesvirus type 4 DNA for nonconserved region F (DMS99 like strain).
DEFINITION 246386
ACCESSION 246386.1 GI:562278
VERSION 246386.1
KEYWORDS Bovine herpesvirus 4.
SOURCE Bovine herpesvirus 4.
ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae.
REFERENCE 1 (Bases 1 to 4179)
AUTHORS Lomonte, P.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1994) Patrick P.L. Lomonte, Virology, University of Liege, Faculty of Veterinary Medicine, Bd de Colonster, Sart Tilman, Liege, 4000, BELGIUM
REFERENCE 2 (Bases 1 to 4179)
AUTHORS Lomonte, P., Bublot, M., van Santen, V., Keil, G.M., Pastoret, P.P. and Thilly, E.
TITLE Analysis of bovine herpesvirus 4 genomic regions located outside the conserved gammaherpesvirus gene blocks
JOURNAL J. Gen. Virol. 76 (Pt 7), 1835-1841 (1995)
FEATURES 97201604
source Location/Qualifiers
 1. 4179
 /organism="Bovine herpesvirus 4"
 /db_xref="taxon:10385"
 /clone="Ecori H (Bublot et al., 1990) tot"
BASE COUNT 1192 a 927 c 951 g 1109 t
ORIGIN

Query Match 41.4%; Score 960; DB 14; Length 4179;
 Best Local Similarity 80.9%; Pred. No. 4.1e-294;
 Matches 1180; Conservative 0; Mismatches 245; Indels 33; Gaps 4;

Qy 507 gaagaagctctgcagctgattactatgtgtgctctggtgtcattatgtctgagcac 566
 Db 2649 GAAGAACCTTTGGCCGGGAGCATCTGTGTGGCCCTTGCTGCTATATGCTGCTGGCGCT 2708
 Qy 567 tgtgctctgaactcttcttcaagttgaagtgtgactcgtacacacttggtctgagtc 626
 Db 2709 TGTTCCTGAGATCTTCCCTCAGGTTTAATGATGATGATGATCTTGGATCTGGAGTC 2768
 Qy 627 caaggaactcnaagccagtaactgtgaatattctgtataattcctgaactccagc 686
 Db 2769 CAGGACTTCCAGAGTACAGACTGTAGGACATGTTTACAAATAGCTGAAGCTGCAGC 2828
 Qy 687 aaagaagctatcaactcgttccaggttcacccgagagggaccagaagacagtgctcagc 746
 Db 2829 AAAGAGATCCATCAACTGTTCTGGATCACGCGAGGGGACCAAGACAGTGTGTCAGGC 2888
 Qy 747 tattctgaataacctgtgaggtcaagaagcgaagccttccacagacccactact 806
 Db 2889 CTTCTTAGACAACTGSAAGTGAAGAAAAGCGGTGCGCTCTCACTGGACCTTACT 2948
 Qy 807 ctccctacacagagactgtgagcaacttcaaggtcgaaggaagtlcataagttccact 866
 Db 2949 CAACATTAACCAAGACTGTGAGCGCTTCAAGGCCCAAGGAAGTTCATACAGTTCACAT 3008
 Qy 867 gaggcaagaagaggtgaggttccctatgcatcactatggtgattatgaagaagatga 926
 Db 3009 GAGTAAAGAGAGCTAATCTCCCATTTGCTTCAATGATGATGATGATGATGATGATGAT 3068
 Qy 927 aaacttgaagagctactgagagctgtgtatgtgtccctcagaacatacactgtcact 986
 Db 3069 AAACCTTGAACGCTGCTGCGAGCGGTGTATGCCCTCAGAACATATACTGTGCCACGT 3128
 Qy 987 ggaatagaagctccacagaacacttcaagaagcggtcgaagaacataattcttgcctcc 1046
 Db 3129 GATGTGAGAGTCCCGACAGACTTTCAAAAGAGCGGTCAAGGCCATATATTTCTCTCC 3188
 Qy 1047 aaatgtctatagccaagttaagctgtgtggtgtgttattagctccctcgtccaggggtga 1106
 Db 3189 CAATGCTTATGAGCCAGTAAATGTGTCGCGGTGTATATGCTCCCTGTCACAGTGA 3248
 Qy 1107 agctgaacctcaactgtcagaagactgtcccaagctcagtgccgtggaataactcct 1166
 Db 3249 AG-----CTGTATGAGAAAGCTGTCTCCAGAGCTCACTAGTGCAGAAAGTACTACT 3298
 Qy 1167 gaalacatgttggagcggacttccctataaagaagaatgcagaagatgttccagctcaca 1226
 Db 3299 GAATACATGCGGGGACAGACTTCCCATTAAGAACCAATGCGAGATGCTCTGCGCTCA 3358
 Qy 1227 gatgttaagtggaagatgcatgagctgagatgagatgctccctcctaagaacaagaagaac 1286
 Db 3359 GATGTGAAGGATGAGAACGTATGAGCTGTGAGTACTTCTGAGTCCCAAAAAATATTCG 3418
 Qy 1287 ctggaataatcaacttgaagtagtgagagacacatcacaactaaccacaagaagaaga 1346
 Db 3419 CTGGAATATACCGCTATAG---GTGACAGACACTGTACCTACACGAAAGATGAAGA 3475
 Qy 1347 tctctccctataatttaactatgatttacaaggaatgctgacatgtgtgctcccgaga 1406
 Db 3476 CCTCCGCCCTGATATATTACCATGTTCAACAGGAATGCTTATTTGTGGCTCTCGAGC 3535
 Qy 1407 ttctgtccacaatgttttgaagaaccttaactcaacaactgattgaaggttaagaaga 1466
 Db 3536 CTTGTGTCACATGCTCTTAAGAACCTTAATCCCAAGAGCTGTTGAATGAGGCAAGA 3595
 Qy 1467 caactatagccagaatgaacactcttggtgcaacctcagaagtgacaggtgagtgctcgtg 1526
 Db 3596 CACCTATAGCCCGACGACAACTCTGGGCTCACCGCTTCAAGCGTCTCTCGTGGATGCTCG 3655

D	660	GTCTATGCTTCCTCGGTCCAGGAGTGACGGGTGAACCTTAACGTGATGGAAGACTGGCTTCAG	719
O	1141	agctcagatgcccgttggaaatactcttcgtgaatacatgttggagcggacttctccataagaagc	1200
D	720	AGCCCCGTGGCCATGGAATATCTCTCGGAACACCTGTGGGACAGACTTTCATCAAAACC	779
O	1201	aatcagaagatggtccaggtctccaagatgtttgaatggggaagaatagatgatgacgaag	1260
D	780	AATGCTGAGATGGTCAAGGCCCTCAAGGCTATTGTAAGAAAGGCGAGAAACAAGTATGGATGCAG	839
O	1261	gtaccctccaagcaacaagaaacccgcctgynaatatcactttgaagtgatgagagaca	1320
D	840	GTACCACTCTCCACTATAAAATATCCCGCTGGAAATATCACTATGAG---GTGACAGACACA	896
O	1321	ttacaataacaaacaagaagaaggtctcccccataacttaacttatctttaaaggg	1380
D	897	TTGCACATGACCAACAGAGAGAGACGCCGCCACTAATTAACCTTAACATGTGTACTCGGG	956
O	1381	aatgcgtacatgtgtgcttcccgagatlttcgtccaacatglttttgaagaaccctaaacc	1440
D	957	AATGCTACATGAGTGGCTTCTCGAGACTTCATTGAACAGTGTTCAGTACATCAAAAGCC	1016
O	1441	caacaactgatatgaatggtgtaaaagaaacttatagcccaagatgaacaccttggccaac	1500
D	1017	CGGCACTGATCGAGTGGTGAATAAGACACTTAATAGTCCCGAAGACACTTTGGGCCACC	1076
O	1501	ctcaagcgtgaacgtttgatgacctgtgctgttcccaacaccccaagtaacacatcca	1560
D	1077	CTCCAGGCTGCTGTGGTAGGCTCGGATCAGATCCCTTGTCATCGAATAATTGACCTGTCA	1136
O	1561	gacatgactcttatgccagcgtgtgtaacgttgcaagtgagatcatatgagggagacacgaatg	1620
D	1137	GACATGAGAGAGCAATTTGGAGACACTTAACCAAGTGGTATGACACATGAGAGACATTGAGAAC	1196
O	1621	ggtgtccctatgtctccctgtctctgtgaacacaaacgggtcatctgtgttataagggct	1680
D	1197	GGGGCACTTAACACGTTCTCTCAGGAAATCCACAGCGGGCTGTGTGCTTATGGGTCA	1256
O	1681	ggggacttgaatgtgatgtcttcaaaaacacatcaactgttggccaacaagtttgaacccaag	1740
D	1257	GGGGCACTGCACTGGATTACTTCAAGAACCAATCACTCTTGGCCAAACAAGTTGACCCAAAG	1316
O	1741	gtagaatataatgctcttgaagtgcttgaagaagaatcactacgttataaaggcatcatagg	1800
D	1317	GTGGATGATATGTTCTTCAAGTGTATTGAAGAAATATTTCGTCAACAAAGCATTATGGG	1376
O	1801	actgaactttgagacaacactatagagcgttctactactgttggggcaagaagacatltacaa	1860
D	1377	ACTGAACATATGAG--ACACTGTAAAGACTTATTGTACTCTGTGTGCAGAAATG-----	1428
O	1861	catgctcaagaactgtctgggacagtgcttgggtggagagaccaggtcttgcgaattgttgca	1920
D	1429	-----CCAGGCACTCACTGGGGGAGTGGGAGTGGGATGATGAGCTGGGGGCTAGACTCCATC	1483
O	1921	tcctttgaataaagagggcgtcattatagatttgggttaagatgacatttgccttgcaaa	1980
D	1484	TCCCTTAGAATGAGGGGGGCTGATATTGAGAACAGGGGTAAGTAGACTTCGCTCTTGGA	1543
O	1961	tttgtctgcctgggtgaatgctgtctgttctctcaaccctaaaccctagtagttctccacta	2040
D	1544	--GGTGTGCGGTGGGAAGGTGGCTATTTCTCCAGCACTCCCTCCCA-----CTTCTTTG	1594
O	2041	acttttcaactaaagtgaagatgaagaactctgtgataagg-----agagtgaaggagat	2096
D	1595	GCTTCTCGCAGACCTTAGAACATGAATAATCTGTACTTGGGTAATTAAGAGGGGGACAT	1654
O	2097	atgtgtgtagagcaacttgattacgtttgaatgtccgtgctgtgtagcttltccatctgtgag	2156
D	1655	GTGACACAGGAGCCCTGGATTTTGGCTGAAGACTCTCTGCAGGTT-----CTTACACAG	1709
O	2157	ctgcgcgttctcaataaatccaggttgtgttgaagcgttggagagagcaactttgatygaagaagaa	2216
D	1710	ACTTTGTTTCAAAATACCTCTGGGGGGGTGTGCTCTGGGTGGGAACTTTATGTGCAAAAGC	1769

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 05:58:21 : Search time 375.65 Seconds
(without alignments)
10599.027 Million cell updates/sec

Title: US-09-874-390-1
Perfect score: 2319
Sequence: 1 ataactgggtttcttatt.....ataataaactcaacaga 2319

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_032802:*

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2319	100.0	2319	21	AAAA8623
2	2205	95.1	2236	22	AAH34463
3	2198.6	94.8	2229	21	AAAG9109
4	2023	87.2	2108	21	AAA96570
5	1990.8	85.8	2128	22	AAAG2134
6	1317	56.8	1317	21	AAA96569
7	1221	52.7	1221	21	AAA96575
8	942.2	40.6	997	19	AAV59800
9	486.2	21.0	777	19	AAV59681

10	456	19.7	524	21	AAA96571	Noncoding region o
11	369.4	15.9	2204	19	AAAX24043	Human core 2GNT DN
12	369.4	15.9	2204	21	AAAC65469	Human heart core 2
13	367.8	15.9	2105	15	AAO61559	cDNA sequence of h
14	364.6	15.7	5010	19	AAAX24042	Rat DHI1 cDNA. Rat
15	364.6	15.7	5010	21	AAAC65468	Diabetic rat heart
16	355	15.3	2102	18	AAT80112	Coding sequence fo
17	284.8	12.3	300	21	AAA01261	Human colon cancer
18	250	10.8	267	21	AAA96572	Noncoding region o
19	199	8.6	997	22	AAAS4282	DNA encoding novel
20	191.8	8.3	1362	22	AAAS0045	Human DNA encoding
21	180.4	7.8	1807	16	AAQ89201	I-branching enzyme
22	180.4	7.8	1807	17	AAT16201	Beta-1,6-N-acetyl
23	180.4	7.8	1807	19	AAV30006	Full length cDNA s
24	180.4	7.8	1807	19	AAV16000	Human beta-1,6-N-a
25	162.8	7.0	378	17	AAT16202	C2GNT catalytic do
26	139.6	6.0	410	22	AAAS4357	DNA encoding novel
27	139.6	6.0	410	22	AAI85732	Human poly nucleoti
28	134	5.8	378	19	AAV30005	cDNA encoding a re
29	133	5.7	210	21	AAAG98742	Human colon cancer
30	129.6	5.6	192	22	AAAG2139	Murine beta-1,6-N-
31	84	3.6	147	22	AAAG2142	Murine beta-1,6-N-
32	69	3.0	69	21	AAAG9574	Core 2 beta-1,6-N-
33	47.2	2.0	387	22	AAAF6413	Novel human polyu
34	46.4	2.0	781	22	AAH98678	Human EST-derived
35	43.6	1.9	66	17	AAT16204	C2GNT C-terminal r
36	43	1.9	99	17	AAT16203	C2GNT C-terminal r
37	39.8	1.7	189	22	ABA65512	Human foetal liver
38	39.8	1.7	189	22	ABA35376	Probe #12042 for g
39	39.8	1.7	189	22	AAK14936	Human brain expres
40	39.8	1.7	189	22	AAK40669	Human bone marrow
41	39.8	1.7	189	22	AAI21432	Probe #11365 for g
42	39.8	1.7	189	22	AAI46721	Probe #15407 used
43	39.8	1.7	189	22	AAI07127	Probe #7118 used t
44	39.8	1.7	434	22	ABA35937	Human foetal liver
45	39.8	1.7	454	22	ABA23688	Probe #2154 for ge

ALIGNMENTS

RESULT 1	
AAA48623	standard; cDNA: 2319 BP.
ID	AAA48623
AC	AAA48623;
XX	
DT	19-SEP-2000 (first entry)
XX	
DE	Human C2/4GNT cDNA.
XX	
KM	Human; C2/4GNT; UDP-N-acetylglucosamine; O-glycan biosynthesis;
KW	O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer; ss.
XX	
OS	Homo sapiens.
XX	
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	496..1812
FT	/*tag= a
FT	/product= "C2/4GNT"
FT	586..605
FT	/*tag= b
FT	1794..1813
FT	/*tag= c
FT	2244..2249
FT	/*tag= d
FT	
PN	WO200034449-A2.
XX	
XX	
XX	15-JUN-2000.
XX	
XX	
PF	03-DEC-1999;
XX	99WO-DK00677.

PR 04-DEC-1998; 98DK-0001605.
 XX
 PA (CLAU/) CLAUSEN H.
 XX Clausen H, Schwientek T;
 XX
 DR WPI; 2000-423407/36.
 DR P-PSDB; AAY94492.
 XX
 XX New nucleic acid molecule encoding UDP-N-acetylglucosamine useful as
 PT probe for the detection of specified glucosaminyltransferase from
 PT other species and related organisms
 PS
 PS Claim 5; Fig 2; 47pp; English.
 XX
 CC The present sequence encodes human UDP-N-acetylglucosamine:
 CC N-acetylglucosamine beta-1,6-N-acetylglucosaminyltransferase
 CC (C2/4GNT). The protein is the third member of the family of O-glycan
 CC beta-1,6-N-acetylglucosaminyltransferases. It adds core 2 or core 4
 CC based O-glycans on to oligosaccharides, glycoproteins and
 CC glycosphingolipids. C2/4GNT can therefore be used in the production of
 CC appropriately glycosylated glycoconjugates with particular enzymatic,
 CC immunogenic, or other biological or physical properties. The nucleotide
 CC sequence is useful as a probe for the detection of C2/4GNT from other
 CC species and related organisms and for the recombinant production of
 CC C2/4GNT polypeptide. The nucleotide sequence was identified by analysis
 CC of EST database sequence information. Oligonucleotides derived from EST
 CC clone 17665 of ATCC were used to isolate two full-length C2/4GNT clones
 CC from a human foreskin genomic PL library by 5' RACE PCR. RT-PCR was
 CC performed using Colo205 human cell line mRNA in order to produce cDNA
 CC for expression of C2/4GNT in Sf9. The control of O-glycan core assembly
 CC has been implicated in tumour progression and metastasis.
 XX
 XX Sequence 2319 BP; 615 A; 528 C; 559 G; 617 T; 0 other;

Query Match 100.0%; Score 2319; DB 21; Length 2319;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 2319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attaacggtttccattatctatcctctgattactctctgtagcagccctc 60
 DB 1 attaacggtttccattatctatcctctgattactctctgtagcagccctc 60
 QY 61 tctcttaagtcacggaagactgctctgctactgtgacactgccccttactcaacatt 120
 DB 61 tctcttaagtcacggaagactgctctgctactgtgacactgccccttactcaacatt 120
 QY 121 ttgtcttggaagccctggaattctgctaatacctacatctgtagtgtgaaggaaac 180
 DB 121 ttgtcttggaagccctggaattctgctaatacctacatctgtagtgtgaaggaaac 180
 QY 181 agatgaagaatctactcaagagcttctctgtaatgatgagaagccagcgagccctg 240
 DB 181 agatgaagaatctactcaagagcttctctgtaatgatgagaagccagcgagccctg 240
 QY 241 caaagatataaagaggagcctgtaaacgttctcttggaacatctataatgacgaaat 300
 DB 241 caaagatataaagaggagcctgtaaacgttctcttggaacatctataatgacgaaat 300
 QY 301 acctttggaaggttagaagatcaggggaatggtgttcaacattctgcccagcgaaca 360
 DB 301 acctttggaaggttagaagatcaggggaatggtgttcaacattctgcccagcgaaca 360
 QY 361 ccggcacttctacttggaacagaaatcagccttgtagaagatcatcccttaagcagga 420
 DB 361 ccggcacttctacttggaacagaaatcagccttgtagaagatcatcccttaagcagga 420
 QY 421 gagaagactactaaagattgtctctcctcaacttcccttgctggtccacccgtc 480
 DB 421 gagaagactactaaagattgtctctcctcaacttcccttgctggtccacccgtc 480
 QY 481 tcccatctctgacgatgtgtcaatggaagagactctgcccagctgcatctactgtggct 540

DB 481 tcccatctctgacgatgtgtcaatggaagagactctgcccagctgcatctactgtggct 540
 QY 541 ctgggcgtcatatgctgcgcgcgcacgtgtgctctgaaacttcttccaggttgaagtc 600
 DB 541 ctgggcgtcatatgctgcgcgcgcacgtgtgctctgaaacttcttccaggttgaagtc 600
 QY 601 gactctgacacacttggtgtgtgagtcagaggaaatctcaaacagcagctctgaaatalc 660
 DB 601 gactctgacacacttggtgtgtgagtcagaggaaatctcaaacagcagctctgaaatalc 660
 QY 661 ttgtataattctctgaaacttccagcaagaagtcatactactgttcaaggtgtaccgga 720
 DB 661 ttgtataattctctgaaacttccagcaagaagtcatactactgttcaaggtgtaccgga 720
 QY 721 ggggaccaaagagagtgcttcaagctatctctgataaactctgagagtcagaagaacgga 780
 DB 721 ggggaccaaagagagtgcttcaagctatctctgataaactctgagagtcagaagaacgga 780
 QY 781 gagccttcaagacacccactactctctcctcaccagagactgtgacacttcaaggtc 840
 DB 781 gagccttcaagacacccactactctctcctcaccagagactgtgacacttcaaggtc 840
 QY 841 gaaaggaagtcatacagttcccaactgtgacgaagaagagtgaggttcctatgtacac 900
 DB 841 gaaaggaagtcatacagttcccaactgtgacgaagaagagtgaggttcctatgtacac 900
 QY 901 tctatgtgtttcaatggaagatgtgaaacttgaagagctgacgtgcagctgtgtatgac 960
 DB 901 tctatgtgtttcaatggaagatgtgaaacttgaagagctgacgtgcagctgtgtatgac 960
 QY 961 cctcagaacatatactgtgtccatgtgtagatgagagtcctccagaacttccaaagagcg 1020
 DB 961 cctcagaacatatactgtgtccatgtgtagatgagagtcctccagaacttccaaagagcg 1020
 QY 1021 gtcaagcaattattcttctgtctcccaatgcttcaatagcagtaagctgtgtcggtg 1080
 DB 1021 gtcaagcaattattcttctgtctcccaatgcttcaatagcagtaagctgtgtcggtg 1080
 QY 1081 gttatgctctctgtgtccaggtgacagctgacactcaactgcatggaagactgtctcag 1140
 DB 1081 gttatgctctctgtgtccaggtgacagctgacactcaactgcatggaagactgtctcag 1140
 QY 1141 agctcagtgccgttggaataacttctcgaatacatgtgtagcagcgttccataaagagc 1200
 DB 1141 agctcagtgccgttggaataacttctcgaatacatgtgtagcagcgttccataaagagc 1200
 QY 1201 aatgcagaagatgtgtccaggtcttcaagaatgttgaatggaggaatacagtgagtcag 1260
 DB 1201 aatgcagaagatgtgtccaggtcttcaagaatgttgaatggaggaatacagtgagtcag 1260
 QY 1261 gtaactcctaaagcaagaagaccgcttggaataatcaacttggagtagtgaagacaca 1320
 DB 1261 gtaactcctaaagcaagaagaccgcttggaataatcaacttggagtagtgaagacaca 1320
 QY 1321 ttacacctaaaccaagaagaagatcctccctataatthaactatgtttacaggg 1380
 DB 1321 ttacacctaaaccaagaagaagatcctccctataatthaactatgtttacaggg 1380
 QY 1381 aatgctgacattgtgtgtctccgagatttgttccaatagttttgaagaaccctaaatcc 1440
 DB 1381 aatgctgacattgtgtgtctccgagatttgttccaatagttttgaagaaccctaaatcc 1440
 QY 1441 caaacaactgttatgtagtgaagaagacactatagcccaatgaacaccccttggccacc 1500
 DB 1441 caaacaactgttatgtagtgaagaagacactatagcccaatgaacaccccttggccacc 1500
 QY 1501 cttaagctgacaggttgatgtcctgtctgttcccaacaccccaagtagacatctca 1560
 DB 1501 cttaagctgacaggttgatgtcctgtctgttcccaacaccccaagtagacatctca 1560
 QY 1561 gacatgactctatctgccaagctgtgtcaagtgtgcagaggtcatgtagggagacatcgataag 1620

Db 1561 gacatgactctatgtccagagctgtgtcaagtggtgaggggtcattagaggaacatgataag 1620
 Oy 1621 ggtgctcttatgtcctctgtcctgtcgaatccaccagcggggtcatctgtgttaagggt 1680
 Db 1621 ggtgctcttatgtcctctgtcctgtcgaatccaccagcggggtcatctgtgttaagggt 1680
 Oy 1681 gggagctgaattgtatgtctcaaacacatcaccgtctgtgccaagaatttaccacaag 1740
 Db 1681 gggagctgaattgtatgtctcaaacacatcaccgtctgtgccaagaatttaccacaag 1740
 Oy 1741 gtagatgaatagtctctcctcagtgcttagaagaatacctacgttataaggccaataggg 1800
 Db 1741 gtagatgaatagtctctcctcagtgcttagaagaatacctacgttataaggccaataggg 1800
 Oy 1801 actgagactttagacacactcttgagagcggtgtgtacccctgtgtggggaagaagatgacaa 1860
 Db 1801 actgagactttagacacactcttgagagcggtgtgtacccctgtgtggggaagaagatgacaa 1860
 Oy 1861 catgctcagaactctgtcggagcagtggtgtggtgagagaccagggcttgcacatctgtgca 1920
 Db 1861 catgctcagaactctgtcggagcagtggtgtggtgagagaccagggcttgcacatctgtgca 1920
 Oy 1921 tcccttaagataagagagcgctctatagatgtgtggttaagtagacatcttgccttgcaa 1980
 Db 1921 tcccttaagataagagagcgctctatagatgtgtggttaagtagacatcttgccttgcaa 1980
 Oy 1981 ttgctgctgggtgtagatgtctgtcttctccaccctcaaccctagtagtcttccacata 2040
 Db 1981 ttgctgctgggtgtagatgtctgtcttctccaccctcaaccctagtagtcttccacata 2040
 Oy 2041 accttctcaactaagtgagaaatgagaaactgtgtgtagaagagagtgagagagagatagt 2100
 Db 2041 accttctcaactaagtgagaaatgagaaactgtgtgtagaagagagtgagagagagatagt 2100
 Oy 2101 ggttagagagacttgatttcaagttgaatgctgtcgtgtgtgaacttcttccatctgtgagctgc 2160
 Db 2101 ggttagagagacttgatttcaagttgaatgctgtcgtgtgtgaacttcttccatctgtgagctgc 2160
 Oy 2161 cgttccataataatccaggttctgtagcgtgtgagagagagacttgtatgtaaaagagacctt 2220
 Db 2161 cgttccataataatccaggttctgtagcgtgtgagagagagacttgtatgtaaaagagacctt 2220
 Oy 2221 ccccttctgtactgttaacttaaaaaataatagctcctgtatccaagaatttacccttactt 2280
 Db 2221 ccccttctgtactgttaacttaaaaaataatagctcctgtatccaagaatttacccttactt 2280
 Oy 2281 ttgtctagtatgtccagaataataataataacttaaacaga 2319
 Db 2281 ttgtctagtatgtccagaataataataataacttaaacaga 2319
 RESULT 2
 AAH34463 standard; cDNA; 2236 BP.
 ID AAH34463
 XX AAH34463:
 AC
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:1545.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; chromosome 15; ss.
 XX
 OS Homo sapiens.
 XX
 FN MO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA.
 XX
 DR WPI: 2001-235357/24.
 DR P-PSDB: AAG75058.
 XX
 PS Claim 1: Page 3190-3191; 9803pp; English.
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytosolic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAG77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SO Sequence 2236 BP; 609 A; 503 C; 548 G; 574 T; 2 other:
 Query Match 95.1%; Score 2205; DB 22; Length 2236;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 64 ctctaaagtcagcgaactgcctctctactgtgtgacgcctgccttaactcaagcattttg 123
 Db 2 ctctaaagtcagcgaactgcctctctactgtgtgacgcctgccttaactcaagcattttg 61
 Oy 124 ttctggagagccttggtatctgtcctaatactatcactgtagtggtctgaaggaaacaga 183
 Db 62 ttctggagagccttggtatctgtcctaatactatcactgtagtggtctgaaggaaacaga 121
 Oy 184 tgaagaacatgacccctcaagagcctcctgtcaatgagagaccagcctgagca 243
 Db 122 tgaagaacatgacccctcaagagcctcctgtcaatgagagaccagcctgagca 181
 Oy 244 agatattaagagagccttgaaactgttctcttggaacattatgataatgcagaaataacc 303
 Db 182 agatattaagagagccttgaaactgttctcttggaacattatgataatgcagaaataacc 241
 Oy 304 ttgtgaggttgaagaatcagggagcattgtgttcacatttgcgccaggaacacgc 363
 Db 242 ttgtgaggttgaagaatcagggagcattgtgttcacatttgcgccaggaacacgc 301
 Oy 364 ccaggtctcacttggaaacagaaatcagccttggaaagatcatcccttaaggcagagag 423
 Db 302 ccaggtctcacttggaaacagaaatcagccttggaaagatcatcccttaaggcagagag 361
 Oy 424 aagctactaaagattgtgtcctcctcacccttccctgtgtcgtccacactgtctcc 483
 Db 362 aagctactaaagattgtgtcctcctcacccttccctgtgtcgtccacactgtctcc 421
 Oy 484 catctgtgacgatgttcaatggaagagactctgacagctgacatttctgtgtgtgtcgt 543
 Db 422 catctgtgacgatgttcaatggaagagactctgacagctgacatttctgtgtgtgtcgt 481

QY	544	ggcgctatactgctgtagccactgtagctctgaaactcttctcaagttgaagtgtgac	603
Db	482	ggctgctatactgctgtagccactgtagctctgaaactcttctcaagttgaagtgtgac	541
QY	604	tctgacaactgggtctcgaggtccagggaaattcccaagccagtaactgtagataatttg	663
Db	542	ctgagccactgtagctgtagtccagggaaattcccaagccagtaactgtagataatttg	601
QY	664	tataattctcgaaactcccgcaaaagatactatacaactgttcagggtccaccgaag	723
Db	602	tataattctcgaaactcccgcaaaaggtctatacaactgttcagggtccaccgaag	661
QY	724	gaccaagagcagtgctccaagctatcttgaataacttgagagtgcaagaagcgaag	783
Db	662	gaccaagagcagtgctccaagctatcttgaataacttgagagtgcaagaagcgaag	721
QY	784	cccttcacaagaccaccatccctccctacacaagagactgtagagaaactcaagctga	843
Db	722	cccttcacaagaccaccatccctccctacacaagagactgtagagaaactcaagctga	781
QY	844	aggaaggttcatacaagttccacctgagcaagaagaggtgagttccctatctcaactc	903
Db	782	aggaaggttcatacaagttccacctgagcaagaagaggtgagttccctatctcaactc	841
QY	904	atgtagtattcalsgaaagattgaaacttgaaaggctactgcgagctgtatagccct	963
Db	842	atgtagtattcalsgaaagattgaaacttgaaaggctactgcgagctgtatagccct	901
QY	964	cagaaacataactctgtccatagttagataaagtcccccagaaacttcaaaagcggtc	1022
Db	902	cagaaacataactctgtccatagttagataaagtcccccagaaacttcaaaagcggtc	961
QY	1024	aaagaacatattctctgctccccaagtgcttcatalagccagtaagctgtgctgggtgt	1083
Db	962	aaagaacatattctctgctccccaagtgcttcatalagccagtaagctgtgctgggtgt	1021
QY	1084	tatgctccctggtccaggggtgcaagctgacctcaactgcaatggaagactgtgccagagc	1144
Db	1022	tatgctccctggtccaggggtgcaagctgacctcaactgcaatggaagactgtgccagagc	1081
QY	1144	tcaagtcgctgtagaaatactctctgtaatacaltgtagagcggaaacttctctaagaagcaat	1201
Db	1082	tcaagtcgctgtagaaatactctctgtaatacaltgtagagcggaaacttctctaagaagcaat	1141
QY	1204	gcaagagctggtccaggtcctcaagaattctaaatgtagagatagacatgtagctcagaagta	1265
Db	1142	gcaagagctggtccaggtcctcaagaattctaaatgtagagatagacatgtagctcagaagta	1201
QY	1264	cctcctaagcacaagaagaaccgctgtgaaatatcaacttgtaggttagttagagacacata	1322
Db	1202	cctcctaagcacaagaagaaccgctgtgaaatatcaacttgtaggttagttagagacacata	1261
QY	1324	caaccacacaacaagaagaagatccctcccccataaattaaactgtgtttacaaggaat	1381
Db	1262	caaccacacaacaagaagaagatccctcccccataaattaaactgtgtttacaaggaat	1321
QY	1384	gctgaacattgtgctctcccgagattctgcacaacatgtttgagaagaccctaattccaa	1443
Db	1322	gctgaacattgtgctctcccgagattctgcacaacatgtttgagaagaccctaattccaa	1381
QY	1444	caactgattgaatgtgtataaagacattatagcccagatgaacactctggtggccacctt	1501
Db	1382	caactgattgaatgtgtataaagacattatagcccagatgaacactctggtggccacctt	1441
QY	1504	cagcgtgacgggtgagtgctggtctgtctccaaaccaccccagaatagacatctccagac	1565
Db	1442	cagcgtgacgggtgagtgctggtctgtctccaaaccaccccagaatagacatctccagac	1501
QY	1564	atgaacttatctgcagaagctgttcaaatgtgcagggtcatatgagggagacaatgcataaggt	1623
Db	1502	atgaacttatctgcagaagctgttcaaatgtgcagggtcatatgagggagacaatgcataaggt	1561
QY	1624	gtctcttatgtccctgctctgtagaattccaccacagcggtactctgtgtttataggggtctgg	1683

Db	1562	gctccttcgtccctcgtcgtcgtgatccaccacgycgcatactcgtttatctatggtgcgtg	1621
Qy	1684	gactctgaatttgatgcttcaaaaacacacacactcttctggcaacaagattttaccacaagta	1743
Db	1622	gactctgaatttgatgcttcaaaaacacacacactcttctggcaacaagattttaccacaagta	1681
Qy	1744	gatgataatgctcttcacgtgctttagaagaatacctacgttataaaggccatactatgagact	1803
Db	1682	gatgataatgctcttcacgtgctttagaagaatacctacgttataaaggccatactatgagact	1741
Qy	1804	gaactcttgagacacacataatgaagagcgttctgctactccttctggggcaagagcattgacaacat	1865
Db	1742	gaactcttgagacacacataatgaagagcgttctgctactccttctggggcaagagcattgacaacat	1801
Qy	1864	gctcgaacctctgctgggagacagcttggtgtggagagaccagggccttgaattctgtgcaccc	1923
Db	1802	gctcgaacctctgctgggagacagcttggtgtggagagaccagggccttgaattctgtgcaccc	1861
Qy	1924	tttagagataaagagagcgtcgtcattagatctgtgggtaagttagatctcttgccttgcaaatg	1983
Db	1862	tttagagataaagagagcgtcgtcattagatctgtgggtaagttagatctcttgccttgcaaatg	1921
Qy	1984	ctgctctggttgtaatgctgcgtctgtctctcaaccctcaaccctgtagttcctccataact	2043
Db	1922	ctgctctggttgtaatgctgcgtctgtctctcaaccctcaaccctgtagttcctccataact	1981
Qy	2044	ttctctacctaagttggaatctgaagaactctgtctgtatagaagagagttgaagagagatctggt	2103
Db	1982	ttctctacctaagttggaatctgaagaactctgtctgtatagaagagagagttgaagagatctggt	2041
Qy	2104	agagaccttgatttccagttgtaatgtaatgctcgtcgtgtagactttccattctgtgagcgtcgt	2163
Db	2042	agagaccttgatttccagttgtaatgtaatgctcgtcgtgtagactttccattctgtgagcgtcgt	2101
Qy	2164	tcctataaatcccaagcttggtagcgtgtggagggagaaactttgatggaagaagaaactctcc	2223
Db	2102	tcctataaatcccaagcttggtagcgtgtggagggagaaactttgatggaagaagaaactctcc	2161
Qy	2224	ttctgtactgttaacttaaaaataatagactcctgatttcaagta	2268
Db	2162	ttctgtactgttaacttaaaaataatagactcctgatttcaagta	2206
RESULT 3			
AAC99109			
ID	AAC99109 standard; cDNA; 2229 BP.		
XX			
AC	AAC99109;		
XX			
XX	09-MAR-2001 (first entry)		
DT			
XX			
DE	Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:337.		
XX			
KW	Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection; diagnosis; identification; cytosolic; neuroprotective; KW		
KW	neurotropic; immunomodulatory; relaxant; contraceptive; gynaecological; KW		
KW	antitumour; cardiac; gene therapy; chromosome mapping; KW		
KW	linkage analysis; tissue identification; tissue typing; forensic; KW		
KW	neural; immune system; muscular; reproductive; gastrointestinal; KW		
KW	pulmonary; cardiovascular; renal; proliferative; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200055320-A1.		
XX			
PD	21-SEP-2000.		
XX			
PF	08-MAR-2000; 2000MO-US05989.		
XX			
PR	12-MAR-1999; 99US-0124270.		
XX			
PA	(HDMA-) HUMAN GENOME SCI INC.		

XX Rosen CA, Ruben SM;
PI
XX
DR WPI: 2000-579444/54.
P-PSDB: AAB54344.
XX
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
PS Claim 1: Page 759-760; 1379pp; English.

CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytosolic,
CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
XX Sequence 2229 BP; 605 A; 501 C; 546 G; 572 T; 5 other;

Query Match 94.8%; Score 2198.6; DB 21; Length 2229;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2198; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 70 gtacagggaactgcccctgctactgtgacctgccccttactcagcagtttgtcttg 129
DB 1 gtcacagggaactgcccctgctactgtgacctgccccttactcagcagtttgtcttg 60
QY 130 gaagccctggtattctgtcataactatactacgttagtgctgaagggaacagatgaaga 189
DB 61 gaagccctggtattctgtcataactatactacgttagtgctgaagggaacagatgaaga 120
QY 190 acatgacctcaaggagcttcctgtcataatagaagaccagcggaccctggcaaaatgt 249
DB 121 acatgacctcaaggagcttcctgtcataatagaagaccagcggaccctggcaaaatgt 180
QY 250 taaaaggagccttgaactgttctccttgacatcttaatagtatgcaaaaataccttttg 309
DB 181 taaaaggagccttgaactgttctccttgacatcttaatagtatgcaaaaataccttttg 240
QY 310 aaggttaagaagatcagggaacatgtgttgcacattgtctgcacaggaaaccggcagtc 369
DB 241 aaggttaagaagatcagggaacatgtgttgcacattgtctgcacaggaaaccggcagtc 300
QY 370 ttcacttggaacacagatacagccttgtaagaagatcacccctaaagcagaagagacta 429
DB 301 ttcacttggaacacagatacagccttgtaagaagatcacccctaaagcagaagagacta 360
QY 430 ctaaaagattgttctcctccacacttccctgtgctcgggtctccacactgtcccaattct 489
DB 361 ctaaaagattgttctcctccacacttccctgtgctcgggtctccacactgtcccaattct 420
QY 490 gtagcagatgttcaatagtagagagactctgcagctgcattactctgtggctctggctgc 549
DB 421 gtagcagatgttcaatagtagagagactctgcagctgcattactctgtggctctggctgc 480

QY 550 tatatgctgtggccactgtggtctgaaacttcttcaggttgagtgactctgac 609
DB 481 tatatgctgtggccactgtggtctgaaacttcttcaggttgagtgactctgac 540
QY 610 cacttggctgtgagtcacagggaatctcaagcagactgttagaataatctgtataat 669
DB 541 cacttggctgtgagtcacagggaatctcaagcagactgttagaataatctgtataat 600
QY 670 ttcttgaacttccagcaagaggtctatcaactgttccagggttccccggaggagccaa 729
DB 601 ttcttgaacttccagcaagaggtctatcaactgttccagggttccccggaggagccaa 660
QY 730 gaggcagctgtcctcaagcttctcgaataacccctgaagtcagaagagcagagccttc 789
DB 661 gaggcagctgtcctcaagcttctcgaataacccctgaagtcagaagagcagagccttc 720
QY 790 acagacaccactactctcctccacagagactgtgagcacttcaaggctgaaagaag 849
DB 721 acagacaccactactctcctccacagagactgtgagcacttcaaggctgaaagaag 780
QY 850 ttcatatcattccactgtgcaagaagagtgaggttccctattgcatatctatgtgtg 909
DB 781 ttcatatcattccactgtgcaagaagagtgaggttccctattgcatatctatgtgtg 840
QY 910 attcatgagaagattgaaaacttgaagcctactgcagctgtgatgcccctcagaac 969
DB 841 attcatgagaagattgaaaacttgaagcctactgcagctgtgatgcccctcagaac 900
QY 970 atatactgtgtcatalgtgtagtgaaggtccccaacttccaagaagcgtgtcaaga 1029
DB 901 atatactgtgtcatalgtgtagtgaaggtccccaacttccaagaagcgtgtgtcaaga 960
QY 1030 atatttctgtctcccaaatgtcttcatagcagcgaagtgtgttggtgtttatgccc 1089
DB 961 atatttctgtctcccaaatgtcttcatagcagcgaagtgtgttggtgtttatgccc 1020
QY 1090 tctgtgtcagaggtgcaagctgacccctcaactgcatgtgaagactgtgtccagagctagtg 1149
DB 1021 tctgtgtcagaggtgcaagctgacccctcaactgcatgtgaagactgtgtccagagctagtg 1080
QY 1150 ccgttgaataacttctctgaatacatatgttggacagacttctcctataaagacaatgcagag 1209
DB 1081 ccgttgaataacttctctgaatacatatgttggacagacttctcctataaagacaatgcagag 1140
QY 1210 atgtgtccaggtcttcaagatgttgaatgggaggaatagcatgtgagtcagaagtactcct 1269
DB 1141 atgtgtccaggtcttcaagatgttgaatgggaggaatagcatgtgagtcagaagtactcct 1200
QY 1270 aagcacaagaagaccgcctgggaataatactacttgaaggttagtgagagacattatcaccta 1329
DB 1201 aagcacaagaagaccgcctgggaataatactacttgaaggttagtgagagacattatcaccta 1260
QY 1330 acccaacaagaagaagatcctcccttataatlaactatgttitaacagggaatgctgtac 1389
DB 1261 acccaacaagaagaagatcctcccttataatlaactatgttitaacagggaatgctgtac 1320
QY 1390 atgtgtgttcccggaatttgcacaaatgttttgaagaaccttaatacccaacaacgt 1449
DB 1321 atgtgtgttcccggaatttgcacaaatgttttgaagaaccttaatacccaacaacgt 1380
QY 1450 attgaaatggtaaaagacacttatagcagagatgaaacactctgggcacaccttcagcgt 1509
DB 1381 attgaaatggtaaaagacacttatagcagagatgaaacactctgggcacaccttcagcgt 1440
QY 1510 gcaagggtgagtgctgtgttcccaaccaccccaagtaacagacatctcagacatgact 1569
DB 1441 gcaagggtgagtgctgtgttcccaaccaccccaagtaacagacatctcagacatgact 1500
QY 1570 tctattgccaagctgtgttcaagtgaggggtcattgagggagatcgtataagggtgtcct 1629
DB 1501 tctattgccaagctgtgttcaagtgaggggtcattgagggagatcgtataagggtgtcct 1560

QY 1630 tatgtccctgctctggaatccaccagcggtactctgcgttattggtggtcgtggacttg 1689
 |||||
 Db 1561 tatgtccctgctctggaatccaccagcggtactctgcgttattggtggtcgtggacttg 1620
 |||||
 QY 1690 aattggtgtcttcaaacatcaacctgttggtccacaagtgtgaaccaagtatgat 1749
 |||||
 Db 1621 aattggtgtcttcaaacatcaacctgttggtccacaagtgtgaaccaagtatgat 1680
 |||||
 QY 1750 aatgctcttaagcttagaagaatacttaagcttataagggcatctatgtagaactt 1809
 |||||
 Db 1681 aatgctcttaagcttagaagaatacttaagcttataagggcatctatgtagaactt 1740
 |||||
 QY 1810 tgagacacactatgagagcgctgtctactcgttggtggaagagcatgtacaaacatgctcag 1869
 |||||
 Db 1741 tgagacacactatgagagcgctgtctactcgttggtggaagagcatgtacaaacatgctcag 1800
 |||||
 QY 1870 aactgtctggacagctgtggtggtggaagacagggcttggcaatcgttggcactcttagg 1929
 |||||
 Db 1801 aactgtctggacagctgtggtggtggaagacagggcttggcaatcgttggcactcttagg 1860
 |||||
 QY 1930 ataagagggcgctctattagattgtgggttaagtagatcttggcttgcgaattgtcgtcct 1989
 |||||
 Db 1861 ataagagggcgctctattagattgtgggttaagtagatcttggcttgcgaattgtcgtcct 1920
 |||||
 QY 1990 ggggtgaatgctgctgtctctccacccctaacccctagtagtctccactaacttcttca 2049
 |||||
 Db 1921 ggggtgaatgctgctgtctctccacccctaacccctagtagtctccactaacttcttca 1980
 |||||
 QY 2050 ctaagtgaagaatgagaactgctgtgatatgagagtgagaagaggaatagtggfagaagca 2109
 |||||
 Db 1981 ctaagtgaagaatgagaactgctgtgatatgagagtgagaagaggaatagtggfagaagca 2040
 |||||
 QY 2110 ctgatttcagtttaaatgacctgctgctgtagacttccattcctgttggagctgctgctctaa 2169
 |||||
 Db 2041 ctgatttcagtttaaatgacctgctgctgtagacttccattcctgttggagctgctgctctaa 2100
 |||||
 QY 2170 taattccaaggttggtagcggtggaaggaactttgatggaagaagaacctccctctgt 2229
 |||||
 Db 2101 taattccaaggttggtagcggtggaaggaactttgatggaagaagaacctccctctgt 2160
 |||||
 QY 2230 actgttaacttaaaaataaataagctcctgattcaaaagta 2268
 |||||
 Db 2161 actgttaacttaaaaataaataagctcctgattcaaaagta 2199
 |||||

RESULT 4
 AAA96570
 ID AAA96570 standard; DNA: 2108 BP.

AC AAA96570;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE A core 2 beta-1,6-N-acetylglucosaminyltransferase DNA fragment.
 XX
 XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GLCNAc-T;
 KM cancer; cardiovascular disorder; inflammatory disorder; asthma;
 KM rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 KM septic shock; adult respiratory distress syndrome; ARDS; cancer;
 KM platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
 KM clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KM diverticulitis; ulcerative colitis; ss.
 XX
 XX Homo sapiens.
 OS
 PN CA2296936-A1.
 XX
 PD 03-AUG-2000.
 XX
 PF 03-FEB-2000; 2000CA-2296936.
 XX
 PR 03-FEB-1999; 99US-0118674.
 XX

PA (GLYC-) GLYCODESIGN INC.
 XX
 PI Koczak B, Lew A;
 XX
 DR WPI; 2000-594746/57.
 XX
 PT New nucleic acid molecules of core 2
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 PT compositions for treatment of disorders mediated by the enzyme
 including cancer, cardiovascular and inflammatory disorders.
 XX
 PS Claim 4; Page 51-52; 66pp; English.
 XX
 CC The present sequence encodes a partial human core 2
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GLCNAc-T) polypeptide.
 CC The polypeptide can be used to treat diseases and disorders, such as
 CC cancer, cardiovascular disorders and inflammatory disorders including
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as
 CC atherosclerosis and clotting can also be treated. The polypeptides of
 CC the invention are predominantly expressed in gastrointestinal tissue
 CC (stomach, colon, intestine, testis) and are elevated in cancer.
 CC Gastrointestinal disorders that may be prevented or treated include
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 CC and ulcerative colitis. The antibodies may be used in
 CC immuno-histochemical analysis, to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular
 CC locations and to quantitate the level of expression.
 XX
 SQ Sequence 2108 BP; 596 A; 464 C; 509 G; 539 T; 0 other;

Query Match 87.2%; Score 2023; DB 21; Length 2108;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2023; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 246 atattaaagagagcctgnaactgttctctggacattatgatgatgcagaataactct 305
 |||||
 Db 18 atattaaagagagcctgnaactgttctctggacattatgatgatgcagaataactct 77
 |||||
 QY 306 ttgagaggttagaagaatcagagggacatggtgttcacatttgcgcgcaggaacagcgc 365
 |||||
 Db 78 ttgagaggttagaagaatcagagggacatggtgttcacatttgcgcgcaggaacagcgc 137
 |||||
 QY 366 agtctcaacttgaaacagaatccacgccttctgaaagatcatccctcaagcagagagaa 425
 |||||
 Db 138 agtctcaacttgaaacagaatccacgccttctgaaagatcatccctcaagcagagagaa 197
 |||||
 QY 426 gctactaaagatgtgtcctcctccacactccctgtgtcgtgtccacactgtctcca 485
 |||||
 Db 198 gctactaaagatgtgtcctcctccacactccctgtgtcgtgtccacactgtctcca 257
 |||||
 QY 486 ttctgtacagatgttcaatggaagaagactctgccaagtcgataacttgggtctggg 545
 |||||
 Db 258 ttctgtacagatgttcaatggaagaagactctgccaagtcgataacttgggtctggg 317
 |||||
 QY 546 ctgtatatagtctggtggcgaactgtggtctgaaacttcttccagttgaagtgtgactc 605
 |||||
 Db 318 ctgtatatagtctggtggcgaactgtggtctgaaacttcttccagttgaagtgtgactc 377
 |||||
 QY 606 tgaccacttgggtctggaagtcacaggaatcctcaagcagactgttagaataatctgtga 665
 |||||
 Db 378 tgaccacttgggtctggaagtcacaggaatcctcaagcagactgttagaataatctgtga 437
 |||||
 QY 666 taatttcctgaactccacagcaagaggtctataactgttcaggggtaccgcagagggga 725
 |||||
 Db 438 taatttcctgaactccacagcaagaggtctataactgttcaggggtaccgcagagggga 497
 |||||
 QY 726 ccaagagcgagtgcttcaaggtatctgataaacttggaaggtcagaagaagcagagagcc 785
 |||||
 Db 498 ccaagagcgagtgcttcaaggtatctgataaacttggaaggtcagaagaagcagagagcc 557
 |||||

QY 786 ttccacagacaccactactcctccctccacagagactgtgagcacttcaaggctgaag 845
 |||||
 Db 558 ttccacagacaccactactcctccctccacagagactgtgagcacttcaaggctgaag 617
 |||||
 QY 846 gaagtctacacgttcccaactgagcaagaaggtgaggttccctattgcatacctat 905
 |||||
 Db 618 gaagtctacacgttcccaactgagcaagaaggtgaggttccctattgcatacctat 677
 |||||
 QY 906 ggtgattctggaagatgtgaagaacttgaaggtctacgtcgaggtgtatgcccctta 965
 |||||
 Db 678 ggtgattctggaagatgtgaagaacttgaaggtctacgtcgaggtgtatgcccctta 737
 |||||
 QY 966 gaacatactgtgtccatgtgtgatagagaagtcgccagaacttcaagaagcggttcaa 1025
 |||||
 Db 738 gaacatactgtgtccatgtgtgatagagaagtcgccagaacttcaagaagcggttcaa 797
 |||||
 QY 1026 agcaattattctgtcttcccaaatgtcttccatagccagtaagcttgcggtgttca 1085
 |||||
 Db 798 agcaattattctgtcttcccaaatgtcttccatagccagtaagcttgcggtgttca 857
 |||||
 QY 1086 tgcctcttggtccagggtgcaagctgcaactgcaactggaagacttctccagagctc 1145
 |||||
 Db 858 tgcctcttggtccagggtgcaagctgcaactgcaactggaagacttctccagagctc 917
 |||||
 QY 1146 agtgcgtggaatactctctgaatacatgtgtgagcgagacttccctataagagcaatgc 1205
 |||||
 Db 918 agtgcgtggaatactctctgaatacatgtgtgagcgagacttccctataagagcaatgc 977
 |||||
 QY 1206 agagatgtgccaggtctcctaagaatgttgaatgggaggaatagcatgagtcagagttacc 1265
 |||||
 Db 978 agagatgtgccaggtctcctaagaatgttgaatgggaggaatagcatgagtcagagttacc 1037
 |||||
 QY 1266 tccctaaagcacaagaagaccgcgtggaatacttcaacttgggtgagtagagacacattaca 1325
 |||||
 Db 1038 tccctaaagcacaagaagaccgcgtggaatacttcaacttgggtgagtagagacacattaca 1097
 |||||
 QY 1326 cctaacaacaagaagaagatctcctccctataatacttaactatgtttacagsgaatgc 1385
 |||||
 Db 1098 cctaacaacaagaagaagatctcctccctataatacttaactatgtttacagsgaatgc 1157
 |||||
 QY 1386 gtacattgtggtctcccgagattcgccacaacatgttttgaagaaccttaatcccaaca 1445
 |||||
 Db 1158 gtacattgtggtctcccgagattcgccacaacatgttttgaagaaccttaatcccaaca 1217
 |||||
 QY 1446 actgattgaatgtgtaaaagacactatagcccgagtgagacacttgggccaaccttca 1505
 |||||
 Db 1218 actgattgaatgtgtaaaagacactatagcccgagtgagacacttgggccaaccttca 1277
 |||||
 QY 1506 gcggtgacggtgagatgctgtgctgttcccaaccaccccaagttacgacatctcagacat 1565
 |||||
 Db 1278 gcggtgacggtgagatgctgtgctgttcccaaccaccccaagttacgacatctcagacat 1337
 |||||
 QY 1566 gactctatctgcaggtctgtcaagtgtgcagaggtcatgagggagacatcgataagggtgc 1625
 |||||
 Db 1338 gactctatctgcaggtctgtcaagtgtgcagaggtcatgagggagacatcgataagggtgc 1397
 |||||
 QY 1626 tccctatgtctcgtctgtggaatccacagcgsggtatctggtttatggggctgggga 1685
 |||||
 Db 1398 tccctatgtctcgtctgtggaatccacagcgsggtatctggtttatggggctgggga 1457
 |||||
 QY 1686 cttgaattgagatgtcaaaacacatcactgtgtgccaacaagttgaccaaaagttaga 1745
 |||||
 Db 1458 cttgaattgagatgtcaaaacacatcactgtgtgccaacaagttgaccaaaagttaga 1517
 |||||
 QY 1746 tgaatagtctctcagtgcttagaagaatacctaagttataaggccatctatgagactga 1805
 |||||
 Db 1518 tgaatagtctctcagtgcttagaagaatacctaagttataaggccatctatgagactga 1577
 |||||
 QY 1806 actttagagacactatgagaggtgtgtctaccgtgtgggcaagagatgttacaacatgc 1865
 |||||
 Db 1578 actttagagacactatgagaggtgtgtctaccgtgtgggcaagagatgttacaacatgc 1637
 |||||
 QY 1866 tcagaactgtctgagacagtggtgtggtgagacagagcttbgcaattcgtgcatcctt 1925
 |||||

Db 1638 tcagaactgtctgagacagtggtgtggtgagacagagcttbgcaattcgtgcatcctt 1697
 |||||
 QY 1926 taggataaagaggctgctttagatgtgtgtggaagaagttcttgccttgcaaatgtct 1985
 |||||
 Db 1698 taggataaagaggctgctttagatgtgtgtggaagaagttcttgccttgcaaatgtct 1757
 |||||
 QY 1986 gctgtggtgaatgctgctgtctctcacccctaacctcagtagttccctccactaactt 2045
 |||||
 Db 1758 gctgtggtgaatgctgctgtctctcacccctaacctcagtagttccctccactaactt 1817
 |||||
 QY 2046 ctccactaaagtgaagaatggaactgctgtgataggagagtggaagggatatagtgttag 2105
 |||||
 Db 1818 ctccactaaagtgaagaatggaactgctgtgataggagagtggaagggatatagtgttag 1877
 |||||
 QY 2106 agcacttgattcagttgaatgctgctgctgtgtgactttccattctgtggaagctgcgcttc 2165
 |||||
 Db 1878 agcacttgattcagttgaatgctgctgctgtgtgactttccattctgtggaagctgcgcttc 1937
 |||||
 QY 2166 ctaataatccaggtttgtagcggtgagagagaacttgatggaagagaaccttccctt 2225
 |||||
 Db 1938 ctaataatccaggtttgtagcggtgagagagaacttgatggaagagaaccttccctt 1997
 |||||
 QY 2226 ctgtactgttaacttaaaataaataagctcctgattcacaagta 2268
 |||||
 Db 1998 ctgtactgttaacttaaaataaataagctcctgattcacaagta 2040
 |||||

RESULT 5
 AAC62134
 ID AAC62134 standard; DNA; 2128 BP.
 XX
 AC AAC62134;
 XX
 DT 06-MAR-2001 (first entry)
 XX
 XX
 DE Nucleotide sequence of beta-1-6-N-acetylglucosaminyltransferase.
 XX
 KW Human; beta-1-6-N-acetylglucosaminyltransferase; CGnT-M; inflammation;
 KW membrane protein; branched sialyl Lex; L-selectin; immune reaction;
 KW inflammation; tissue rejection; tumour metastasis; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key location/Qualifiers
 FT CDS 354..1670
 FT FT /*tag= a
 FT FT /product= "beta-1-6-N-acetylglucosaminyltransferase"
 FT FT 2100..2105
 FT FT /*tag= b
 XX
 PN US6136580-A.
 XX
 PD 24-OCT-2000.
 XX
 PF 19-JAN-1999; 99US-0233506.
 XX
 PR 19-JAN-1999; 99US-0233506.
 XX
 PA (BURN-) BURNHAM INST.
 XX
 PI Fukuda M, Yeh J;
 XX
 DR WPI: 2001-040238/05.
 DR P-PSDB: AAB30518.
 XX
 PT New CGnT-M polypeptides having core 2, core 4 and I branching
 PT beta-1-6-N-acetylglucosaminyltransferase activities for preparing
 PT reagents useful for diagnosing, preventing or treating inflammation or
 PT tumour metastasis -
 XX
 PS Example 1; Fig 4; 25pp; English.
 XX

CC The present sequence encodes a human
CC beta-1-6-N-acetylglucosaminyltransferase polypeptide, with core2, core4
CC and I branching activities. It is designated C2GnT-M. C2GnT-M is a
CC membrane protein that is predominantly expressed in colon, small
CC intestine, trachea, stomach and thyroid, as well as in certain cancer
CC cell lines. C2GnT-M polypeptides may be used to prepare molecules having
CC highly branched sialyl Lex and L-selectins, which may be subsequently
CC used to modulate immune reactions, e.g. inflammation and tissue
CC rejection, and to prevent or inhibit tumour metastasis.
XX
S0 Sequence 2128 BP; 569 A; 477 C; 526 G; 556 T; 0 other;

Query Match 85.8%; Score 1990.8; DB 22; Length 2128;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2004; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY	244	agatatgaagagagcgtgaactgttctcttggaacatctatgaagtcaagaataacc	303
DB	102	agatatgaagagagcgtgaactgttctcttggaacatctatgaagtcaagaataacc	161
QY	304	ttttgagaggttaagaatcaggggacatgtgttcacatttgctgcacagaaacacg	363
DB	162	ttttgagaggttaagaatcaggggacatgtgtgttcacatttgctgcacagaaacacg	221
QY	364	ccagcttctcacttggaaacagaatacgcctgtgtgaagagatcatccctaaagcagagag	423
DB	222	ccagcttctcacttggaaacagaatacgcctgtgtgaagagatcatccctaaagcagagag	281
QY	424	aagctactaaaggattgtctctctcctccaccttccctgtgtgtgtccacccgtctcc	483
DB	282	aagctactaaaggattgtctctctcctccaccttccctgtgtgtgtccacccgtctcc	341
QY	484	cattctgtgaagatgtgttcataatgtgaagagactctgcacagctgcatcttactgttggtctg	543
DB	342	cattctgtgaagatgtgtgttcataatgtgaagagactctgcacagctgcatcttactgttggtctg	401
QY	544	ggctgcataatgtcgtgcgcacatgtgtgtcttgaaactttcttccaggttgaaagtgtgac	603
DB	402	ggctgcataatgtcgtgcgcacatgtgtgtcttgaaactttcttccaggttgaaagtgtgac	461
QY	604	tcctgacacacttgggtctctggagttccagaggaactccaagaacagctactgttagaataatctg	663
DB	462	tcctgacacacttgggtctctggagttccagaggaactccaagaacagctactgttagaataatctg	521
QY	664	tataattctctgaacttccagcaagaaggtctatacactgttcaagggttaccaccgaggg	723
DB	522	tataattctctgaacttccagcaagaaggtctatacactgttcaagggttaccaccgaggg	581
QY	724	gaaccaagaagcagtgcttcaaggtctatcttgaaataacctggaagtcagaagaacgagag	783
DB	582	gaaccaagaagcagtgcttcaaggtctatcttgaaataacctggaagtcagaagaacgagag	641
QY	784	ccttccagagacccactactcttccctccacagagacttggagactttaaagctgaa	843
DB	642	ccttccagagacccactactcttccctccacagagacttggagactttaaagctgaa	701
QY	844	aggaagtctacatagttccacatgacgaagaagaggttgaggtccctatgcatct	903
DB	702	aggaagtctacatagttccacatgacgaagaagaggttgaggtccctatgcatct	761
QY	904	atggttgatctagaagaagtgtgaaacttgaagaagctactgcagctgtgtatgcccct	963
DB	762	atggttgatctagaagaagtgtgaaacttgaagaagctactgcagctgtgtatgcccct	821
QY	964	cagaacatatactgtgtccatgagtgtgagagagtcctccagagaacttccaagaagcggttc	1023
DB	822	cagaacatatactgtgtccatgagtgtgagagagtcctccagagaacttccaagaagcggttc	881
QY	1024	aaagaacattatcttctgcttcccaatgtcttcatagccaagttaagcgtgttcggtgtt	1083
DB	882	aaagaacattatcttctgcttcccaatgtcttcatagccaagttaagcgtgttcggtgtt	941

QY	1084	tatgctctctgttccagaggttgacaagctgacctcaactgcatgtgaagaactgtctccagagc	1143
DB	942	tatgctctctgttccagaggttgacaagctgacctcaactgcatgtgaagaactgtctccagagc	1001
QY	1144	tcagtgccgrrggaatactctctctgaatacatgttgaggcggagacttctctataaagaagaat	1203
DB	1002	tcagtgccgrrggaatactctctctgaatacatgttgaggcggagacttctctataaagaagaat	1061
QY	1204	gcagaagatgttccagagcttccaaagtgtgaatgtggaggaatagcagaggtccagagta	1263
DB	1062	gcagaagatgttccagagcttccaaagtgtgaatgtggaggaatagcagaggtccagagta	1121
QY	1264	cctcctaagcaacaagaaccccgctggaataatatactttaggttagtgagagacacatta	1323
DB	1122	cctcctaagcaacaagaaccccgctggaataatatactttaggttagtgagagacacatta	1181
QY	1324	caacctaaaccaagaagaagatccctcccttatataatttaactatagtgttacagggat	1383
DB	1182	caacctaaaccaagaagaagatccctcccttatataatttaactatagtgttacagggat	1241
QY	1384	gcgtacattgtgtgtcccgagatttgcaccaatgttttgaagaacccctaaatcccaa	1443
DB	1242	gcgtacattgtgtgtgtcccgagatttgcaccaatgttttgaagaacccctaaatcccaa	1301
QY	1444	caactgtatgaatgtgttaaaagacactataagcccaagtatacacctctgtggccaacct	1503
DB	1302	caactgtatgaatgtgttaaaagacactataagcccaagtatacacctctgtggccaacct	1361
QY	1504	cagcgtgcacggttgatgtgcctgtctgttcccaaccaccccaagtcagacatctcagac	1563
DB	1362	cagcgtgcacggttgatgtgcctgtctgttcccaaccaccccaagtcagacatctcagac	1421
QY	1564	atgacttctatgtccagagctgtgtcaagttgagcaaggtcatgagggagacatcgaataaggt	1623
DB	1422	atgacttctatgtccagagctgtgtcaagttgagcaaggtcatgagggagacatcgaataaggt	1481
QY	1624	gtctccttatgtctccctgtctctgtgaaatccacacaggggtatctgtgtttatggggctcggg	1683
DB	1482	gtctccttatgtctccctgtctctgtgaaatccacacaggggtatctgtgtttatggggctcggg	1541
QY	1684	gacttgaaatgtgaatgtgttccaaacacacacactgtgttgccaaacaagtttgaccaccaagta	1743
DB	1542	gacttgaaatgtgaatgtgttccaaacacacacactgtgttgccaaacaagtttgaccaccaagta	1601
QY	1744	gatgataatgtctcttccagtgctttagaagaatactactagttataaaggccatctatgagact	1803
DB	1602	gatgataatgtctcttccagtgctttagaagaatactactagttataaaggccatctatgagact	1661
QY	1804	gaactttagagacacactatgagagcgttgtctactgtgtggggccaagacatgatacaaat	1863
DB	1662	gaactttagagacacactatgagagcgttgtctactgtgtggggccaagacatgatacaaat	1721
QY	1864	gctcagaactgtctggagacagttgtgggttgagagacagggcttttgcaattcgtgtgaccc	1923
DB	1722	gctcagaactgtctggagacagttgtgggttgagagacagggcttttgcaattcgtgtgaccc	1781
QY	1924	tttagagaataaggggtctgataatgagatgtgttgagtaagtaatcttctgcttggaaatg	1983
DB	1782	tttagagaataaggggtctgataatgagatgtgttgagtaagtaatcttctgcttggaaatg	1841
QY	1984	ctgcctgtgtgtaatgtcgtctgtctctccacccctaaacctagttagttctctccactaact	2043
DB	1842	ctgcctgtgtgtaatgtcgtctgtctctccacccctaaacctagttagttctctccactaact	1901
QY	2044	tttctcactaaggttagaataagtaagttgtgtgataagggagaggtgaaggagataatgtgt	2103
DB	1902	tttctcactaaggttagaataagtaagttgtgtgataagggagaggtgaaggagataatgtgt	1961
QY	2104	agaagcattgatttcaattgaatgaatgcctgcctgtgtagacttccattcctgtgagactgcggt	2163
DB	1962	agaagcattgatttcaattgaatgaatgcctgcctgtgtagacttccattcctgtgagactgcggt	2021
QY	2164	tcctaataattccaggtttgtgtagcgttgagaggaactttagatgagaaagaaccttccc	2223

Accession	Year	Accession	Year
D0	2022	ttccaaatattccaggtttctgtacgttgagagaaacttgaatggaagagac--ttcc	2079
Oy	2224	ttctgtactgttaacttaaaaaataa	2251
D0	2080	ttctgtactgttaacttaaaaaataa	2107

RESULT 6

AAA96569
ID AAA96569 standard; DNA; 1317 BP.

AC AAA96569;

DT 08-FEB-2001 (first entry)

DE DNA encoding a core 2 beta-1,6-N-acetylglucosaminyltransferase.

KW Core 2 eta-1,6-n-acetylglucosaminyltransferase; core2b GICMNC-;
KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis; ss.

OS Homo sapiens.

FH	key	Location/Qualifiers
----	-----	---------------------

FT	CDS	1..1317
----	-----	---------

```
ET
    /product= "core 2
```

PN CA2296936-A1.

PD 03-AUG-2000.

PF 03-FEB-2000; 2000CA-2296936.

PR 03-FEB-1999; 99US-0118674.

PA (GLYC-) GLYCODESIGN INC.

PI Korczak B, Lew A;

DR WPI: 2000-594746/57.

XX

PT New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
XX including cancer, cardiovascular and inflammatory disorders.
PS
Claim 4; Page 50; 66pp; English.

Claim 4; Page 50; 66pp; English.

CC The present sequence encodes a human core 2

CC The polypeptide can be used to treat diseases and disorders, such as

CC asthma, rheumatoid arthritis, inflammatory bowel disease,

CC (ARDS) and cancer. Various platelet-mediated pathologies such as

CC the invention are predominantly expressed in gastrointestinal

CC Gastrointestinal disorders that may be prevented or treated include

CC and ulcerative colitis. The antibodies may be used in

CC localize it to particular cells and tissues and to specific subcellular

XX

50 Sequence 1317 BP; 359 A; 307 C; 322 G; 329 T; 0 other;

Query Match 56.88; Score 1317; DB 21; Length 1317;

```

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1317; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

OY	496	atbtttcaatgaaagaaactctgcgaactgacattactgtgaggtctctggcgtcataag	555
Db	1	atggttcaatgaaagaaactctgcgaactgacattactgtgaggtctctggcgtcataag	60
OY	556	ctgcgtgcgaactgtggtctctgaaacttctcttcacagttgaaagtgtgacctgtgacactgtg	615
Db	61	ctgcgtgcgaactgtggtctctgaaacttctcttcacagttgaaagtgtgacctgtgacactgtg	120
OY	616	ggtctgaggtctccagggaaattctcaagcgaactactgtgaggaatatctctgtataatttctgt	675
Db	121	ggtctgaggtctccagggaaattctcaagcgaactactgtgaggaatatctctgtataatttctgt	180
OY	676	aaacttccagaagaagaggtctatacactgtctcaggggtctcccccgaaggagccaagaagatga	735
Db	181	aaacttccagaagaagaggtctatacactgtctcaggggtctcccccgaaggagccaagaagatga	240
OY	736	gtgcttcaaggtatctctgaaataactgtgaggtgtaagaagaagcgaagaccttccacagac	795
Db	241	gtgcttcaaggtatctctgaaataactgtgaggtgtaagaagaagcgaagaccttccacagac	300
OY	796	aaccaactaacctctcccacacagaagactgtgagcaacttcaaggtctgaaaggaaattcata	855
Db	301	aaccaactaacctctcccacacagaagactgtgagcaacttcaaggtctgaaaggaaattcata	360
OY	856	caagtctccactgtgacaaagaagaggtgaggtttccctatgtgacatactctatgtgttcaat	915
Db	361	caagtctccactgtgacaaagaagaggtgaggtttccctatgtgacatactctatgtgttcaat	420
OY	916	gagaaagattgaaaaacttggaaaggtctactctgcggtgtgtatgtcccttcagaacatatac	975
Db	421	gagaaagattgaaaaacttggaaaggtctactctgcggtgtgtgtatgtcccttcagaacatatac	480
OY	976	tgttctcaattgtgatactgaaaggtctccccaagaacttccaagaagggcgtgtccaagaacataat	1035
Db	481	tgttctcaattgtgatactgaaaggtctccccaagaacttccaagaagggcgtgtccaagaacataat	540
OY	1036	tcttgcgtctcccaaatgtcttctcaatagccagtaagctgtgtccgggtgtgttatagtctctccgtg	1095
Db	541	tcttgcgtctcccaaatgtcttctcaatagccagtaagctgtgtccgggtgtgttatagtctctccgtg	600
OY	1096	tccaggggtgcaagctctgacccccaactactgcatgtgaagactgtgtccagaaggtcgaatgcgcgtg	1155
Db	601	tccaggggtgcaagctctgacccccaactactgcatgtgaagactgtgtccagaaggtcgaatgcgcgtg	660
OY	1156	aaatacttctctgaaatactgtggaacgagacttctctcataaagaagcaatgcagaagatgtgtc	1215
Db	661	aaatacttctctgaaatactgtggaacgagacttctctcataaagaagcaatgcagaagatgtgtc	720
OY	1216	caggtctcctaagatgttggaaatggaagaatagaaatggaatcagaaggttactctcttaagac	1275
Db	721	caggtctcctaagatgttggaaatggaagaatagaaatggaatcagaaggttactctcttaagac	780
OY	1276	aaagaaaacccgtggaataatcaacttggagtgtagagagacacatttaaaccttaaacac	1335
Db	781	aaagaaaacccgtggaataatcaacttggagtgtagagagacacatttaaaccttaaacac	840
OY	1336	aagaagaagaagatctctcccccattataatttaactacatagtttctacagggaaatgcgttaacttgtg	1395
Db	841	aagaagaagaagatctctcccccattataatttaactacatagtttctacagggaaatgcgttaacttgtg	900
OY	1396	gtcttcccaagatttcgtcccaacatgttttgaagaacccccaatcccaacaactgtatga	1455
Db	901	gtcttcccaagatttcgtcccaacatgttttgaagaacccccaatcccaacaactgtatga	960
OY	1456	tgggttaaaagacattatagccagaatgaaacagctctgggcaaccccttaacgctgtacaggtg	1515

Db	961	tgggttaaaagacacttatagccagatgtgaacacctcttggtgccacccttcagcgtgcaag	1020
QY	1516	tgaatgcctgagctctgttcccaaccccccagaatcagacatctcagacatgaacttcatt	1575
Db	1021	tgaatgccttgagctctgttcccaaccccccagaatcagacatctcagacatgaacttcatt	1080
QY	1576	gcacagctgtgtcgaatgagcaggggtcactgtagggagacatcgtcaaaaggctgccttatgct	1635
Db	1081	gcacagctgtgtcgaatgagcaggggtcactgtagggagacatcgtcaaaaggctgccttatgct	1140
QY	1636	ccctgcctcgtgaatccaacccagcgagctatcgtcgtttalagggctctggagacttgaaatgg	1695
Db	1141	ccctgcctcgtgaatccaacccagcgagctatcgtcgtttalagggctctggagacttgaaatgg	1200
QY	1696	atgcttcaaaacatcacctcgttgcgcacaacagtttgacccaagaagtatgatatgct	1755
Db	1201	atgcttcaaaacatcacctcgttgcgcacaacagtttgacccaagaagtatgatatgct	1260
QY	1756	cttcagctgtctagaagaatacctaagttataagggccatctalaggaactgaacttga	1812
Db	1261	cttcagctgtctagaagaatacctaagttataagggccatctalaggaactgaacttga	1317

CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC Gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
SQ Sequence 1221 BP: 342 A; 285 C; 295 G; 299 T; 0 other;

Query Match 52.7%; Score 1221; DB 21; Length 1221;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 ttgaagtgaactcgcgaaccttggtcgtgcgagtcacggaatccaaagcagtaactgt 651
|||||
DB 1 ttgaagtgcgacctgcgacaccttgcgtgcgagtcacggaatccaaagcagtaactgt 60
|||||
QY 652 aggaatatcttgataattcctgcgaacttcgaagaaagagtcatacaactgttcaagg 711
|||||

RESULT 7
 ID AAA96575 standard; DNA; 1221 BP.
 AC AAA96575;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region DNA.
 XX
 XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
 KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
 KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
 KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KW diverticulitis; ulcerative colitis; ss.
 XX
 OS Homo sapiens.
 XX
 PN CA2296936-A1.
 PD 03-AUG-2000.
 XX
 PF 03-FEB-2000; 2000CA-2296936.
 XX
 PR 03-FEB-1999; 99US-0118674.
 XX
 PA (GLXC-) GLYCODESIGN INC.
 XX
 PI Korczak B, Lew A;
 DR WPI; 2000-594746/57.
 XX
 PT New nucleic acid molecules of core 2
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 PT compositions for treatment of disorders mediated by the enzyme
 PT including cancer, cardiovascular and inflammatory disorders.
 XX
 PS Claim 4; Page 54; 66pp; English.
 XX
 CC The present sequence encodes a fragment of a human core 2
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide
 CC The polypeptide can be used to treat diseases and disorders, such as
 CC cancer, cardiovascular disorders and inflammatory disorders including
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as
 CC atherosclerosis and clotting can also be treated. The polypeptides of
 CC the invention are predominantly expressed in gastrointestinal tissue

QY	712	gtcacccgaaggagaccagaagacatgctctcaagctatcttgaataacttgtagtcaag	771
Db	121	gtcacccgaaggagaccagaagacatgctctcaagctatcttgaataacttgtagtcaag	180
QY	772	aagaagcgagagccttccacagaccaccactacctctccctccacagagactgtagcac	831
Db	181	aagaagcgagagccttccacagaccaccactacctctccctccacagagactgtagcac	240
QY	832	ttcaagcgctgaagaagatcacaagctccactgagcaagaagaaggtgtagtccct	891
Db	241	ttcaagcgctgaagaagatcacaagctccactgagcaagaagaaggtgtagtccct	300
QY	892	attgatactcctatgtagtatcatatagaaagattggaaaacttggaaaggtactgcagct	951
Db	301	attgatactcctatgtagtatcatatagaaagattggaaaacttggaaaggtactgcagct	360
QY	952	gtgtagtccctccacagaacacatactatgtagtgcacatgtagtagaagatccccaagaatttc	1011
Db	361	gtgtagtccctccacagaacacatactatgtagtgcacatgtagtagaagatccccaagaatttc	420
QY	1012	aaagagcggtgataaaggcaattatcttctgctctcccaaatgcttccatagccagtagctg	1071
Db	421	aaagagcggtgataaaggcaattatcttctgctctcccaaatgcttccatagccagtagctg	480
QY	1072	gttcgggttggtttaaagctccctccgtgctccaaagggtgcacaagctcgaactcgaactgtagaag	1131
Db	481	gttcgggttggtttaaagctccctccgtgctccaaagggtgcacaagctcgaactcgaactgtagaag	540
QY	1132	ttgctccagagctcagtcagtcgctggaataactctccctggaataactgtggcgagacttctct	1191
Db	541	ttgctccagagctcagtcagtcgctggaataactctccctggaataactgtggcgagacttctct	600
QY	1192	ataaagagcaaatgcacagagatgtgtccaaagctcccaaatgttgaatggagggagaaagacatg	1251
Db	601	ataaagagcaaatgcacagagatgtgtccaaagctcccaaatgttgaatggagggagaaagacatg	660
QY	1252	gagtcagagtagtctctcctaagacacaaagaaacccgctggaataatcatcttgaagttagtg	1311
Db	661	gagtcagagtagtctctcctaagacacaaagaaacccgctggaataatcatcttgaagttagtg	720
QY	1312	agagacaaattcacctaacaacaagaagaaggtctccctcccttatataatttaactatg	1371
Db	721	agagacaaattcacctaacaacaagaagaaggtctccctcccttatataatttaactatg	780
QY	1372	tttaacaggaatctgctacatgttgctctcccgagaaattgtctccaacatgttttgaagaac	1431
Db	781	tttaacaggaatctgctacatgttgctctcccgagaaattgtctccaacatgttttgaagaac	840
QY	1432	cctaatacccacaacatgttagttagtgggttaaagacacttatagccagatgaacactc	1491

Db 841 cctaaccacaacatgatitgaatggtaaagaacacttatagccagatgaacacc 900
QY 1492 tgggcacccttcacgcygcagcgtgatgcctgtctgttcccaaccaccagttac 1551
Db 901 tgggcacccttcacgcygcagcgtgatgcctgtctgttcccaaccaccagttac 960
QY 1552 gacatctcagacatgactctatctgcagcgtgtcgaattgagcaggaagac 1611
Db 961 gacatctcagacatgactctatctgcagcgtgtcgaattgagcaggaagac 1020
QY 1612 atcgataaagggtcctctatgtctccctgtctgtgaatccacagcggtatctgcgt 1671
Db 1021 atcgataaagggtcctctatgtctccctgtctgtgaatccacagcggtatctgcgt 1080
QY 1672 tatgggcttgggacttgatgtgatgtgtcttcaaacatcactcttggccaacagtt 1731
Db 1081 tatgggcttgggacttgatgtgatgtgtcttcaaacatcactcttggccaacagtt 1140
QY 1732 gaccacaaggtagatgataatgctctcagtgcttagaagaatacctatgataagac 1791
Db 1141 gaccacaaggtagatgataatgctctcagtgcttagaagaatacctatgataagac 1200
QY 1792 atctatgggactgaacttga 1812
Db 1201 atctatgggactgaacttga 1221

RESULT 8
AAVS9800
ID AAVS9800 standard: DNA; 997 BP.
AC AAVS9800;
XX
DT 19-JAN-1999 (first entry)
XX
DE Human secreted protein gene 171 clone HTEDJ34.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN MO9839448-A2.
XX
PD 11-SEP-1998.
XX
PF 06-MAR-1998; 98WO-US04493.
XX
PR 02-OCT-1997; 97US-0061060.
PR 07-MAR-1997; 97US-0038621.
PR 07-MAR-1997; 97US-0040161.
PR 07-MAR-1997; 97US-0040162.
PR 07-MAR-1997; 97US-0040163.
PR 07-MAR-1997; 97US-0040333.
PR 07-MAR-1997; 97US-0040336.
PR 07-MAR-1997; 97US-0040626.
PR 11-APR-1997; 97US-0043311.
PR 11-APR-1997; 97US-0043312.
PR 11-APR-1997; 97US-0043313.
PR 11-APR-1997; 97US-0043314.
PR 11-APR-1997; 97US-0043568.
PR 11-APR-1997; 97US-0043569.
PR 11-APR-1997; 97US-0043576.
PR 11-APR-1997; 97US-0043578.
PR 11-APR-1997; 97US-0043580.
PR 11-APR-1997; 97US-0043669.

PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 23-MAY-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0044500.
PR 23-MAY-1997; 97US-0044501.
PR 23-MAY-1997; 97US-0044502.
PR 23-MAY-1997; 97US-0044503.
PR 23-MAY-1997; 97US-0044581.
PR 23-MAY-1997; 97US-0044582.
PR 23-MAY-1997; 97US-0044583.
PR 23-MAY-1997; 97US-0044584.
PR 23-MAY-1997; 97US-0044585.
PR 23-MAY-1997; 97US-0044586.
PR 23-MAY-1997; 97US-0044587.
PR 23-MAY-1997; 97US-0044588.
PR 23-MAY-1997; 97US-0044589.
PR 23-MAY-1997; 97US-0044590.
PR 23-MAY-1997; 97US-0044592.
PR 23-MAY-1997; 97US-0044593.
PR 23-MAY-1997; 97US-0044594.
PR 23-MAY-1997; 97US-0044595.
PR 23-MAY-1997; 97US-0044596.
PR 23-MAY-1997; 97US-0044597.
PR 23-MAY-1997; 97US-0044598.
PR 23-MAY-1997; 97US-0044599.
PR 23-MAY-1997; 97US-0044600.
PR 23-MAY-1997; 97US-0044601.
PR 23-MAY-1997; 97US-0044612.
PR 23-MAY-1997; 97US-0044613.
PR 23-MAY-1997; 97US-0044614.
PR 23-MAY-1997; 97US-0044615.
PR 23-MAY-1997; 97US-0044616.
PR 23-MAY-1997; 97US-0044618.
PR 23-MAY-1997; 97US-0044633.
PR 06-JUN-1997; 97US-0044634.
PR 06-JUN-1997; 97US-0044635.
PR 13-JUN-1997; 97US-0044610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0051928.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.

PR 07-MAR-1997; 97US-0040162.
PR 07-MAR-1997; 97US-0040163.
PR 07-MAR-1997; 97US-0040333.
PR 07-MAR-1997; 97US-0040334.
PR 07-MAR-1997; 97US-0040336.
PR 07-MAR-1997; 97US-0040626.
PR 11-APR-1997; 97US-0043311.
PR 11-APR-1997; 97US-0043312.
PR 11-APR-1997; 97US-0043313.
PR 11-APR-1997; 97US-0043314.
PR 11-APR-1997; 97US-0043566.
PR 11-APR-1997; 97US-0043567.
PR 11-APR-1997; 97US-0043576.
PR 11-APR-1997; 97US-0043578.
PR 11-APR-1997; 97US-0043580.
PR 11-APR-1997; 97US-0043669.
PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0047492.
PR 23-MAY-1997; 97US-0047500.
PR 23-MAY-1997; 97US-0047501.
PR 23-MAY-1997; 97US-0047502.
PR 23-MAY-1997; 97US-0047503.
PR 23-MAY-1997; 97US-0047581.
PR 23-MAY-1997; 97US-0047582.
PR 23-MAY-1997; 97US-0047583.
PR 23-MAY-1997; 97US-0047584.
PR 23-MAY-1997; 97US-0047585.
PR 23-MAY-1997; 97US-0047586.
PR 23-MAY-1997; 97US-0047587.
PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 06-JUN-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 13-JUN-1997; 97US-0048974.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056300.
PR 22-AUG-1997; 97US-0056331.
PR 22-AUG-1997; 97US-0056332.
PR 22-AUG-1997; 97US-0056336.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.

PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057669.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.

(HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
PI Feng P, Ferlie AM, Fischer CU, Florence KA, Greene JM, Hu JS;
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX
XX WPI: 1998-506364/43.
DR P-PSDB: AAW74898.
XX
XX
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
XX
XX Claim 1: Page 410-411; 721pp; English.
XX
XX This sequence represents a nucleic acid molecule designated Gene 171
CC from the human cDNA clone hREDJ34 (deposited as clone ATCC 97904 and
CC ATCC 209050) which encodes a secreted human protein. The gene can be
CC used to generate fusion proteins by linking to the gene to a human
CC immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of
CC the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 186
CC polynucleotides, based on which tissues they are most highly expressed in
XX (see AAV59511 for described uses).
XX
XX Sequence 777 BP; 203 A; 142 C; 194 G; 198 T; 40 other;
SO

Query Match 21.0%; Score 486.2; DB 19; Length 777;
Best Local Similarity 81.9%; Pred. No. 2,3e-145;
Matches 610; Conservative 34; Mismatches 67; Indels 34; Gaps 7;

QY 1558 tcagacatgactctatctatccagagctggtcgaagtgatcagaggaagacatgat 1617
DB 8 kcgacatgactctatctatccagagctggtcgaagtgatcagaggaagacatgat 67
QY 1618 aaggggtcctctatgctccctgctctggaaccacacagggatcgtcgtttatgg 1677
DB 68 aaggggtcctctatgctccctgctctggaaccacacagggatcgtcgtttatgg 127
QY 1678 gctgggacttgatgattgctcaaacacatcacctlttgccaagaattgacca 1737

```
Dh 128 gctgggactagaaattgagatctcaaaaacacatcactgltggcacaagttagacca 187
Qy 1738 aaggtagatgataatgctctcaagtcttagaagaatactagcttataagccc-----at 1793
Dh 188 aaggtagatgataatgctctcaagtcttagaagaatactagcttataagggccattct 247
Qy 1794 ctatggactgaactttagacacatat-----gagagcggtgtgctacc 1837
Dh 248 attggaactgaactttagaagacacamatltgagaagcggtgtgctaccggtgtgggggc 307
Qy 1838 tgtgggcaagaagcatctacaaatctcagaaactgtgtggac--agtgtgggtggga 1895
Dh 308 caaagagcaatgtaccaaacaatggyycargaaattggykkggammcarxxxxkkgkkgga 367
Qy 1896 gaccagggccttgcaa--ctcgtggacatcctta--gataaagaagcgctctattaagat 1950
Dh 368 rrmcmrggytgcgaawttskkgcmwccytttaggttaarrggckltwatlgat 427
Qy 1951 tgtgggta---agtagatcttctgcttgcaaatgtgctgctgggtgaatgtgctgt 2006
Dh 428 tgtgggtaagtagatcttctgcttgcaaatgtgctgctgggtgaatgtgctgt 487
Qy 2007 tctctc---acccttaaccctagtagtccctcctaacttctcactaagttagaaga 2063
Dh 488 tctctcmaccctcaaccctagtagtccctcctaacttctcactaagttagaaga 547
Qy 2064 gaactgtctatagaggagagtgaaaggagatgtgttagagcaactgttcaagtg 2123
Dh 548 gaactgtctatagaggagagtgaaaggagatgtgttagagcaactgttcaagtg 607
Qy 2124 aatgcctgctgtgacttcttccattctgtgagctgctgccttccataatccaggttg 2183
Dh 608 aatgcctgctgtgacttcttccattctgtgagctgctgccttccataatccaggttg 667
Qy 2184 gtagcgtggaagagaactttagtgaagaagaaccttcccttctgtaactgttaactaa 2243
Dh 668 gtagcgtggaagagaactttagtgaagaagaaccttcccttctgtaactgttaactaa 727
Qy 2244 aataatagctcctgattcaagaata 2268
Dh 728 aataatagctcctgattcaagaata 752

RESULT 10
AAA96571
ID AAA96571 standard; DNA; 524 BP.
XX
AC AAA96571;
XX
DT 08-FEB-2001 (first entry)
XX
DE Noncoding region of core 2 beta-1,6-N-acetylglucosaminyltransferase gene.
XX
KW Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GLCNAC-T;
KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis; ss.
XX
OS Homo sapiens.
XX
PN CA2296936-A1.
XX
PD 03-AUG-2000.
XX
PF 03-FEB-2000; 2000CA-2296936.
XX
PR 03-FEB-1999; 99US-0118674.
XX
PA (GLYC-) GLYCODESIGN INC.
XX
```

```
PI Korczak B, Lew A;
XX
DR WPI; 2000-594746/57.
XX
PT New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
PS Disclosure; Page 52; 66pp; English.
XX
CC The present sequence represents a noncoding region from a human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GLCNAC-T) gene.
CC The polypeptide can be used to treat diseases and disorders, such as
CC cancer, cardiovascular disorders and inflammatory disorders including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC Gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
SQ Sequence 524 BP; 171 A; 90 C; 127 G; 136 T; 0 other;
```

```
Query Match 19.7%; Score 456; DB 21; Length 524;
Best Local Similarity 100.0%; Pred. No. 9,5e-136;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1813 gacacataagagagcgttctgactctgtgggcaagaagcatgatacaaatgctcagaac 1872
Dh 1 gacacacataagagagcgttctgactctgtgggcaagaagcatgatacaaatgctcagaac 60
Qy 1873 ttgtctggagacagtggtgtgggagacagagcgttgaatctgtgagcatcctttagata 1932
Dh 61 ttgtctggagacagtggtgtgggagacagagcgttgaatctgtgagcatcctttagata 120
Qy 1933 agagagcctgctatagattgtgggtaagtagatcttctgcttgcgaattgctcggg 1992
Dh 121 agagagcctgctatagattgtgggtaagtagatcttctgcttgcgaattgctcggg 180
Qy 1993 tgaatgcctgtgttcttccacccttaaccctagtagtcttccctcaacttctcacta 2052
Dh 181 tgaatgcctgtgttcttccacccttaaccctagtagtcttccctcaacttctcacta 240
Qy 2053 agtgaagatgagaactgctgtgataaggagagtgaaaggagagatgtgtgtagaacact 2112
Dh 241 agtgaagatgagaactgctgtgataaggagagtgaaaggagagatgtgtgtagaacact 300
Qy 2113 gattcagtgtaagcctgctgtgagcttccattctgtgagacgctcgcttctaata 2172
Dh 301 gattcagtgtaagcctgctgtgagcttccattctgtgagacgctcgcttctaata 360
Qy 2173 ttcaagtttgtagcgtgtgggaggaacttgaaggaaaggaaaccttcccttctgact 2232
Dh 361 ttcaagtttgtagcgtgtgggaggaacttgaaggaaaggaaaccttcccttctgact 420
Qy 2233 gttacttaaaaaataatagctcctgattcaagaata 2268
Dh 421 gttacttaaaaaataatagctcctgattcaagaata 456
```

```
RESULT 11
AA24043
ID AA24043 standard; cDNA; 2204 BP.
XX
AC AA24043;
```


KW Human; diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;
 KW UDP-GlcNAc:Galbeta1-3GalNAc:alpha6-beta1-6
 KW N-acetylglucosaminyltransferase; ss.
 XX Homo sapiens.
 XX US6131578-A.
 XX 17-OCT-2000.
 XX 02-OCT-1997; 97US-0943058.
 XX 02-OCT-1996; 96US-0046876.
 XX (KING/) KING G L.
 XX (NISH/) NISHIO Y.
 XX (KOYA/) KOYA D.
 XX (DENN/) DENNIS J W.
 XX (WARR/) WARREN C E.
 XX Nishio Y, Koya D, King GL, Warren CE, Dennis JW;
 PI MPI; 2000-678642/66.
 DR P-PSDB; AAB30298.
 XX Preventing or treating cardiomyopathy associated with diabetes mellitus
 PT and hyperglycemia, comprises administering a substance that inhibits
 PT core 2 N-acetylglucosamine-T activity -
 XX
 XX Claim 3; Fig 9; 21pp; English.
 PS
 CC The present invention describes the human UDP-GlcNAc:Galbeta1-3GalNAc
 CC alpha6-beta1-6 N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein
 CC and coding sequences. The enzyme is associated with cardiomyopathy in
 CC diabetes and hyperglycaemia sufferers. The invention also provides
 CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used
 CC in treatment.
 XX
 XX Sequence 2204 BP; 641 A; 414 C; 498 G; 651 T; 0 other;
 SQ

Query Match 15.9%; Score 369.4; DB 21; Length 2204;
 Best Local Similarity 60.2%; Pred. NO. 1.7e-107;
 Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

QY 772 aagaagcgaagacgttcacagacacccactacccctccacagagagctgagcac 831
 DB 490 aaaaagcgccctcggtgacacccgacgactatataaacgtgacagtgctcttc 549
 QY 832 ttcaagcgtgaagaagctacacagttccactgagcaagaagagtggttcctc 891
 DB 550 ttcatcaagagacgcaaatatattgtagaaccccttagtaagaagagcgagttcca 609
 QY 892 attgacactatggtgatcatgagagaattgaaaactttggaagcgactcgagct 951
 DB 610 atagcatattctaaagtggtcaccacaaagatggaatgcttgacagcgctgagggcc 669
 QY 952 gtgtatgccctcagaacataactgtgtccatgtgagatgagaagtcgccagaacttc 1011
 DB 670 atcatatgctcgaattcttctgctcatgtagtgcacaaatcgcgagatctctat 729
 QY 1012 aaagagcggtcacaagaattattctgctcccaaatgtctcatagccagtaagctg 1071
 DB 730 ttgactgcaagtatggtcactgcctctgtttagtaagtctttgtgycacagcgattg 789
 QY 1072 gtccgggtgtttatgctctcgtgtcagagtggaagctgaactcaactctgagaagac 1131
 DB 790 gagagtggtgttatgcatcgtgagcggtgtcagcgtgacccactctgcatgaagat 849
 QY 1132 ttgctcagagcgtcagcgctggaataactctcgtatacatctgtagcgagacttcct 1191
 DB 850 ctctatgcaatgagtgcaactggaagtaactgtgataaactcttggtatgagatttccc 909

QY 1192 ataaagcaatgcaagatggtccaggtctccaagatgttgatggtggaatgacatg 1251
 DB 910 attaaacacactagaattgttgcagaagctcaagtgtttaaattggaagaacaactg 969
 QY 1252 gattcagaggttaactccttaagcacaagaagaccgcctcggaatataactttgggtatg 1311
 DB 970 gaaacggaagagatgcatcccataaagaagaaggggaagaagcgtatgaggtcgt 1029
 QY 1312 agagacacattacacccaacaagaagaagatccctccctataatttaattatg 1371
 DB 1030 aatggaagagctgacaacaacagggactgcaaaaatgtcttccatccggaacaactc 1089
 QY 1372 tttaacaggaatgctgacattgtgcttccgagatttgcaccaactgtttgaagaac 1431
 DB 1090 ttcttgcagctgctcactcttcgtgtcagtaggagatgttggttatgtactacaagat 1449
 QY 1432 cctaatacccacaactgattggaatggtaagaacacttatagccagatggaacactc 1491
 DB 1150 gaaaaaacccaagaagtgagagtggtgcacaagacacatacagccttgatgatactg 1209
 QY 1492 tgggcacaccttcaagcgtgacagtggtgactgtctgttcccaaccacccaagttac 1551
 DB 1210 tgggcacacatccaagaagattcctggaagctccgggtcactccctgcagcacaagat 1269
 QY 1552 gacatctcaacatgactttattgccaagctggtcaagtggaaggtcattgaggaagac 1611
 DB 1270 gatctatctgacatgcaagcagtggtccaggtttgtcaatgtagactttgaggtgat 1329
 QY 1612 atgataaggggtctctctatgctccctgtcctggaatccacacagcggtctatcggt 1671
 DB 1330 gtttccaaggtgtctccctccctcccgccctgcgagtaggaatcattgtgctcagtgatc 1389
 QY 1672 tatgggctgtggagactgtaattgtagtcttcaaaacacatcactgttgccaacaagtt 1731
 DB 1390 ttcgagctgtgtgactgtaactgtagtctgctgcaaacacactgtttgccaataagtt 1449
 QY 1732 gaccacaaggtatgataatgtcttctcagtgctttagaagaataactcagttataagc 1790
 DB 1450 gacgtgtagtgtgacactcttgcacatccagttgttgatgagatgtgagacacaagc 1508

RESULT 13
 AA061559 standard; cDNA; 2105 BP.
 ID AA061559
 XX AA061559;
 XX 07-OCT-1994 (first entry)
 DE cDNA sequence of human core 2 beta 1-6
 DE N-acetylglucosaminyltransferase (C2Gnt or core 1-6 AGT).
 KW C2Gnt; 1-6 AGT; core 2 beta 1-6 n-acetylglucosaminyltransferase;
 KW O-glycan; ss.
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 220..1504
 FT /*tag= a
 FT polyA_signal 1913..1918
 FT /*tag= b
 FT /*tag= d
 FT /*tag= 248..314
 FT /*tag= c
 FT /*tag= signal/membrane-anchoring domain
 XX BP590747-A.
 XX 06-APR-1994.
 XX
 PD 29-SEP-1993; 93EP-0250268.
 XX
 PF 01-OCT-1992; 92US-0955041.
 PR

XX This invention describes a method for screening for a substance that
 CC may be used to prevent or treat cardiomyopathy associated with diabetes
 CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
 CC (UDP-GlcNAc:Galbeta1-3GalNAc-alpha beta-1,6-N-acetylglucosaminyl-
 CC transferase) with an acceptor substrate and a sugar nucleotide donor in
 CC the presence of a test substance under conditions whereby the core 2
 CC GlcNAc-T produces a reaction product, determining the amount of reaction
 CC product, and comparing the amount of reaction product with the amount
 CC obtained in the absence of the test substance, where lower amounts of
 CC reaction product in the presence of the test substance indicate that the
 CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)
 CC methods for preventing or treating cardiomyopathy associated with
 CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
 CC germ cells and somatic cells all contain a DNA construct introduced into
 CC the animal or an ancestor of the animal at an embryonic stage, where
 CC incorporation of the DNA construct into the germ line of the animal
 CC causes the animal to develop cardiomyopathy similar to that associated
 CC with diabetes mellitus and hyperglycaemia. This sequence encodes the rat
 CC DH1 protein which is used to describe the method of the invention.
 XX

SQ Sequence 5010 BP; 1369 A; 987 C; 1093 G; 1561 T; 0 other;

Query Match 15.78; Score 364.6; DB 19; Length 5010;
 Best Local Similarity 58.3%; Pred. No. 1e-105;
 Matches 659; Conservative 0; Mismatches 469; Indels 3; Gaps 1;

QY 686 caaagaggtatcatcaactgttcagggtcccgaggagcccaagagcggtgtcagg 745
 Db 959 ccataagcaatgttaattgaccacaaagttttacagggtgaccgccgaagaattccagaag 1018
 QY 746 ctattctgaataaccttggaggtcaa--gagaagcgaagcgctttccacagacaccat 802
 Db 1019 tgaagcttggatgatactcaacagtgcaattcaagaagcgctccgagcgaccccgatgact 1078
 QY 803 accttccctcaccagaagactgtgagcaacttcaagcgtgaaggaagtatacagttcc 862
 Db 1079 atataaacatgacccggagactgcgctctctcatcaggaacacgaatatatttgagc 1138
 QY 863 cactgagcaagaagaggtgaggttcccttgcataccttaigtatcatgaagaga 922
 Db 1139 cccttactaagaagaaggttggcttccaatgtcatalccatgtgttccatcataaga 1198
 QY 923 ttgaanaacttgaagactactcgaagctgttgtatgcccctcacaataatactgttcc 982
 Db 1199 ttgacatgcttgaacagctcctcctgagggccatcatalgcacagaattctcactgcatc 1258
 QY 983 atgttgatgagaaggtcccccgaacttcaagaagcggtccaaagcaatattcttgc 1042
 Db 1259 acgttgacagaagaagcagaagatccttlttagccggtgcagggcatlgtcatcctgct 1318
 QY 1043 tcccaaatgtctcatagccagtaagctgtgttcgggtgtttagctccctcggtccagg 1102
 Db 1319 ttgataatgtcttctgtgcccagccagcttgaagagtggtttatataagctcccgaggtc 1378
 QY 1103 tgaagctgacactcaactgcatgaagaactgtgtccacagctcagtgccgttgaataact 1162
 Db 1379 ttaaggtgcctcaactgatgaagactgttacagaatgaatgacaaactgtgagact 1438
 QY 1163 tcttgaatacatgttggaaggaacttccctataaagacaaatgcagagatgtccagctc 1222
 Db 1439 tgatcaactcttctgtgtatgtattcccttataaaccaacctgtgaattgttcagagc 1498
 QY 1223 tcaagaatgttggaatggagaatagcatgtagctcagaggtaccccttaagcacaaagaa 1282
 Db 1499 tcaagtccttcacagaggaacagcctggaagaacatgagaatgtccccaacaagaa 1558
 QY 1283 ccgcgttgaataatcacttggatgtagagagacacattacacctaacaagaaga 1342
 Db 1559 agaggttgtaaaaaacgatacaggtgtgtggaacggaagcgcagacaacacgtgagttgca 1618
 QY 1343 aggatctcccttataattactatgtttacaggaatgcgtacatgttgctcc 1402

Db 1619 aggcgcagctccactcaaaactcctctcttccaagcagcgctatttctgtgtaacta 1678
 QY 1403 gaeatttgcacacatgttttgaagacccttaatcccaacactgattgaatggtaa 1462
 Db 1679 gggaattgtgaagctatgtcttgaataataataatattcaaaagttcatgtaatggcac 1738
 QY 1463 aagacattatagcccaagatgaacacactcgtggccaccccttcacggtgcaggtgatagc 1522
 Db 1739 agcacatacagcccgatgagttcctctggccaccatccaagaatccctgaagtc 1798
 QY 1523 ctgcttctgttcccaaccaccacaagtagaacatcttaagcagactctatgtccagc 1582
 Db 1799 ctgcttctccctcaagcacaatgaatgacttctgtgacatgaatgctgtgactagt 1858
 QY 1583 tggctcaagtgccaggggtcagagggaagacatcgaataaggtgtccctatgtccctgct 1642
 Db 1859 ttgtcaagtgacataacttgaagggcagatgttcccaatggcgcccttaaccacgtgca 1918
 QY 1643 ctggaatccacacagcggtatctgcgttatagtgggtggtggaactgtaattgatactc 1702
 Db 1919 gtgagtcacatgtgcgctctgtgtgctcttggagtgtgtgacttgagctgagtgctgc 1978
 QY 1703 aaaaacatcacctgttggccacaagaatttgaaccacaagtgatgataatgctctcagt 1762
 Db 1979 gcaaacaccacttcttgcacaataagtttgcatagtgagtgtgaccccttccctcag 2038
 QY 1763 gcttaagagatcctcactagcttaagagccatccatcagtggaactggaacttgg 1813
 Db 2039 gtttgaagaacatctgaggcataaagccctggaagacctaagaactgaag 2089

RESULT 15
 AAC65468
 ID AAC65468 strand; cDNA; 5010 BP.
 XX
 AC AAC65468;
 XX
 DT 12-FEB-2001 (first entry)
 XX
 DE Diabetic rat heart core 2 GlcNAc-T coding sequence.
 XX
 KW Rat; diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;
 KW UDP-GlcNAc:Galbeta1-3GalNAc:alpha beta-1,6-N-acetylglucosaminyltransferase; ss.
 KW
 OS Rattus sp.
 XX
 PN US6131578-A.
 XX
 PD 17-OCT-2000.
 XX
 PF 02-OCT-1997; 97US-0943058.
 XX
 PR 02-OCT-1996; 96US-0046876.
 XX
 PA (KING/) KING G L.
 PA (NISH/) NISHIO Y.
 PA (KOYA/) KOYA D.
 PA (DENN/) DENNIS J W.
 PA (WARR/) WARREN C E.
 XX
 PI Nishio Y, Koya D, King GL, Warren CE, Dennis JW;
 XX
 DR WPI: 2000-678642/66.
 DR P-PSDB; AAB30297.
 XX
 PT Preventing or treating cardiomyopathy associated with diabetes mellitus
 PT and hyperglycaemia, comprises administering a substance that inhibits
 PT core 2 N-acetylglucosamine-T activity -
 XX
 PS Example 1; Fig 3A; 21pp; English.
 XX

CC The present invention describes the human UDP-GlcNAc:Galbeta1-3galNAc
CC alphaL beta1-6-N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein
CC and coding sequences. The enzyme is associated with cardiomyopathy in
CC diabetes and hyperglycemia sufferers. The invention also provides
CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used
CC in treatment.

XX Sequence 5010 BP; 1369 A; 987 C; 1093 G; 1561 T; 0 other;

Query Match 15.7%; Score 364.6; DB 21; Length 5010;
Best Local Similarity 58.3%; Pred. No. 1e-105;
Matches 659; Conservative 0; Mismatches 469; Indels 3; Gaps 1;

QY 686 caaagaggtctatcacaggttcagggtgcacccggaggagcccaagagaggttcagg 745
DB 999 ccaataagcaatgttaactgcacaaagtttacaagggtgacccagaagaatccgaag 1018
QY 746 ctattctgaataacctgtagtcaaa---gaagaagcgagagccttcacaagacccact 802
DB 1039 tgaagcttgagatctaacaagtcgaatcaagaagcgctccggagcgacacccgact 1078
QY 803 acctctccctccacagagactgtgagcaactcaaggctgaagaagtlcatacagttcc 862
DB 1079 atataaacatgacccggactgcgctcctcatcagacacgcgaatatatatgagc 1138
QY 863 cactgagcaagaagagtgtaggtccctattgcatactctatgtgtatctatgagaaga 922
DB 1139 cccttaccacaagaagaagttgtgtcttcaatgcataatcacaatgtgtcatalaaga 1198
QY 923 ttgaaacttgaagaagctactgcagctgtgtatgcccctcagacacatatgtgtcc 982
DB 1199 ttgacatgttgacaggtcctcgaggccatctatagccaagaatttctactgcatc 1258
QY 983 atgttgatgagagagtcgccagaaacttccaagaagcggtcaagaacatttcttgc 1042
DB 1259 acgttgacagaaaagcagagagatcctttagccggtgacaggtcattgcatctgc 1318
QY 1043 tcccaaatgtcttcaagccatgaagctgttcggtgtgttatgctcctcgtccaggg 1102
DB 1319 ttgataaatgtcttcttgccagccaggttcggaagtggtgtatacgcctcccgag 1378
QY 1103 tgaagctgcagctcaactgcagtcagactgttccagagctcagtgccgttgaaatct 1162
DB 1379 ttaagctgcagctcaactgcagtcagactgttacagaatgaaatgcagtcgaag 1438
QY 1163 tccgttaatacatgttgagcgagcttctctataaagaacatgcagaatgtccaggtc 1222
DB 1439 tgaataatcttgtgtatgtatcttccatlaaaacacacctggaattgtcaggaag 1498
QY 1223 tcaagaatttgatggaggaatagcatgaggtcagaggttaaccttaagccaagaaga 1282
DB 1499 tcaagctctccacaggggaaaacacgctcgaaactggaagatgctccccaagaaga 1558
QY 1283 ccggttggaatatcaactttaggtagtgagagacacatlaacctaaaccaagaaga 1342
DB 1559 agaggttggaaaaaaacgatacaaggttgtagcggaagctgaacaacactgtgactca 1618
QY 1343 aggatctccctctataaattaactatgtttacaggaatgtcgtacattgtgcttcc 1402
DB 1619 aggcgagcctcactcaaaactcctctcttcttcagcgagcgtcatttgcgtcacta 1678
QY 1403 gagattcgtccaataatgttttgaaagaaaccttaatcccaactgtatgtggttaa 1462
DB 1679 gggaaatagtgtatgtgtcgtgaataataaatalatcaaaagtlcaggaatggtgcac 1738
QY 1463 aagacattatagccacagatgacacctctggccaccccttgcagctgacaggtgagtc 1522
DB 1739 aggaacatacagcccgatgtgtcctctgtggccaacatccaagaagatccctgaagtc 1798
QY 1523 ctggtctgttcccaacaccccaagtlacgacatcagacatcgtacttatgtccaggc 1582
DB 1799 ctggtctctctccctcaagcctaagtatgactgtctgacatgaaatgtctgctgtagt 1858

QY 1583 tggtaagrtggcaggggtcaagggagacatcgataaagggtgtccttatgtcctgct 1642
DB 1859 ttgtcaatgagcaataacttcgaaggcgatgttccaatggtgcgctctatccacagtgca 1918
QY 1643 ctggaatccaccagcggtatctcggttataatgggtctggggagacttgaattgactcttc 1702
DB 1919 gtgaggtccatgtgtgctctgtgtgcttcttgaggttggtgacttgaactgtgatgtgc 1978
QY 1703 aaacatcaactgttggccaagaagttgaccaccaaggtagatgataatgtcctcagt 1762
DB 1979 gaaacaccacttcttgccaataagttgacatgatgtgactcccttgccctcaggt 2038
QY 1763 gcttgaagaataactagttataagggccatctatggagactgaacttgag 1813
DB 2039 gtttggaagaacatctgagcataaagccctggagaccttgaacgttaag 2089

Search completed: September 26, 2002, 08:19:19
Job time: 8458 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 04:26:55 ; Search time 72.39 seconds
(without alignments)
7868.823 Million cell updates/sec

Title: US-09-874-390-1

Perfect score: 2319

Sequence: 1 ataactgggtttccatctt.....ataataatacaacaga 2319

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTOS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1990.8	85.8	2128	3	US-09-233-506-1
2	369.4	15.9	2105	1	US-07-955-041-3
3	369.4	15.9	2105	1	US-08-227-455-3
4	369.4	15.9	2105	1	US-08-472-482-3
5	369.4	15.9	2105	1	US-08-487-069-3
6	355	15.3	2102	3	US-09-063-237-3
7	180.4	7.8	1807	1	US-08-118-906-13
8	180.4	7.8	1807	1	US-08-486-196-13
9	180.4	7.8	1807	1	US-08-488-135-13
10	180.4	7.8	1807	2	US-08-474-065-13
11	162.8	7.0	378	1	US-08-118-906-3
12	162.8	7.0	378	1	US-08-486-196-3
13	162.8	7.0	378	1	US-08-488-135-3
14	162.8	7.0	378	2	US-08-474-065-3
15	134	5.8	378	1	US-08-118-906-1
16	134	5.8	378	1	US-08-486-196-1
17	134	5.8	378	1	US-08-488-135-1
18	134	5.8	378	2	US-08-474-065-1
19	129.6	5.6	192	3	US-09-233-506-9
20	84	3.6	147	3	US-09-233-506-13
21	43.6	1.9	66	1	US-08-118-906-11
22	43.6	1.9	66	1	US-08-486-196-11
23	43.6	1.9	66	1	US-08-488-135-11
24	43.6	1.9	66	2	US-08-474-065-11
25	43	1.9	99	1	US-08-118-906-7
26	43	1.9	99	1	US-08-486-196-7
27	43	1.9	99	1	US-08-488-135-7

28	43	1.9	99	2	US-08-474-065-7	Sequence 7, Appl1
c 29	35.4	1.5	7218	1	US-08-232-463-14	Sequence 14, Appl1
30	35	1.5	99	1	US-08-118-906-5	Sequence 5, Appl1
31	35	1.5	99	1	US-08-486-196-5	Sequence 5, Appl1
32	35	1.5	99	1	US-08-488-135-5	Sequence 5, Appl1
33	35	1.5	99	2	US-08-474-065-5	Sequence 5, Appl1
c 34	33.2	1.4	1941	4	US-09-402-002-1	Sequence 1, Appl1
c 35	33	1.4	11283	2	US-08-603-753D-3	Sequence 3, Appl1
c 36	33	1.4	11283	3	US-09-099-753-3	Sequence 3, Appl1
c 37	33	1.4	11283	4	US-08-986-106-3	Sequence 3, Appl1
38	32.8	1.4	246240	2	US-08-724-394A-20	Sequence 20, Appl1
39	32.8	1.4	246240	2	US-08-724-394A-21	Sequence 21, Appl1
40	32.8	1.4	246240	2	US-08-724-394A-22	Sequence 22, Appl1
41	32.4	1.4	7218	1	US-08-232-463-14	Sequence 14, Appl1
42	32.2	1.4	3164	1	US-07-723-002C-3	Sequence 3, Appl1
43	32.2	1.4	5194	4	US-08-642-846-1	Sequence 1, Appl1
44	32.2	1.4	5194	4	US-09-264-604-1	Sequence 1, Appl1
45	31.8	1.4	900	1	US-08-218-026-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-233-506-1
Sequence 1, Application US/09233506
Patent No. 6,165,580

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru

APPLICANT: Yeh, Jinn-Chern

TITLE OF INVENTION: A Beta-1-6-N-Acetylglucosaminyltransferase That Forms

TITLE OF INVENTION: Core 2, Core 4 and I Branches

FILE REFERENCE: P-LI 3415

CURRENT APPLICATION NUMBER: US/09/233,506

CURRENT FILING NUMBER: 09/09233506

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 2128

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (354)..(1670)

US-09-233-506-1

Query Match 85.8%; Score 1990.8; DB 3; Length 2128;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2004; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY	244	agatataaagggaagcgcgaacactgttcttgaacatcttaagaatgtcagaataacc	303
DB	102	agatataaagggaagcgcgaacactgttcttgaacatcttaagaatgtcagaataacc	161
QY	304	tttggagggttaagaagatcgaaggaatgtgttcacattgttcgccaaggaacacg	363
DB	162	tttggagggttaagaagatcgaaggaatgtgttcacattgttcgccaaggaacacg	221
QY	364	ccagcttcaacttggaagaacgaatcagcgttggaagaatcatccctaagaagagag	423
DB	222	ccagcttcaacttggaagaacgaatcagcgttggaagaatcatccctaagaagagag	281
QY	424	aagctactaaagatgtgtctccctccacactcctgtgtcgtcgtccacactgtctcc	483
DB	282	aagctactaaagatgtgtctccctccacactcctgtgtcgtcgtcgtccacactgtctcc	341
QY	484	cattctgacatgtgtcacaatggaagagacctcgcagctcattactgtggtctcg	543
DB	342	cattctgacatgtgtcacaatggaagagacctcgcagctcattactgtggtctcg	401
QY	544	ggctgcatatgctgctgccaactgtgctctgaacattcttcagggttgaagtgcac	603
DB			

Dh 402 ggcgtctatatgctgctgcccactgctgctctgaaactttctcttcagggttgagtgac 461
Qy 604 tctgaccacttggtctgagctccagggaatctcaagccagtaacttgaagaatctctg 663
Dh 462 tctgaccacttggtctgagctccagggaatctcaagccagtaacttgaagaatctctg 521
Qy 664 tataattctctgaacttccagcaagaagtgctatacaactgttcaagggttccagcga 723
Dh 522 tataattctctgaacttccagcaagaagtgctatacaactgttcaagggttccagcga 581
Qy 724 gaccagaaggaagtgctgctgagctatctctgaataacttgagtgcaagaagaagcga 783
Dh 582 gaccagaaggaagtgctgctgagctatctctgaataacttgagtgcaagaagaagcga 641
Qy 784 ccttccacagaccaccactactctccctccacagagacttgagcacttcaagcgtgaa 843
Dh 642 ccttccacagaccaccactactctccctccacagagacttgagcacttcaagcgtgaa 701
Qy 844 aggaagttcatacagttcccaactgagcaagaaggttgaggttccctatgtactact 903
Dh 702 aggaagttcatacagttcccaactgagcaagaaggttgaggttccctatgtactact 761
Qy 904 atggttgattataggaagatgtaaaactttgaaagctactgagcagctgtgtacgccc 963
Dh 762 atggttgattataggaagatgtaaaactttgaaagctactgagcagctgtgtacgccc 821
Qy 964 cagaacataactgtgtcactgtgtgagatgagagtcaccagaaactttcaagaagcgctc 1023
Dh 822 cagaacataactgtgtcactgtgtgagatgagagtcaccagaaactttcaagaagcgctc 881
Qy 1024 aaagaacttatttctgcttcccaaatgtcttcaatagcagtaagctgtgtcgggtgtc 1083
Dh 882 aaagaacttatttctgcttcccaaatgtcttcaatagcagtaagctgtgtcgggtgtc 941
Qy 1084 tatccccctggtctcaggtgtcaggtcagctcagctcagctcagctcagctcagctcagc 1143
Dh 942 tatccccctggtctcaggtgtcaggtcagctcagctcagctcagctcagctcagctcagc 1001
Qy 1144 tcagtcgcgtgagaaactctctctgaatacatgttgagagcggaacttccataaagaacat 1203
Dh 1002 tcagtcgcgtgagaaactctctctgaatacatgttgagagcggaacttccataaagaacat 1061
Qy 1204 gcagagatgtgtccaggtctctcaagatgttgatgagaggaataacatgagatcagagta 1263
Dh 1062 gcagagatgtgtccaggtctctcaagatgttgatgagaggaataacatgagatcagagta 1121
Qy 1264 cctcctaagcaagaagaacccgctgtgaaatatacaacttggagtgagagacacatta 1323
Dh 1122 cctcctaagcaagaagaacccgctgtgaaatatacaacttggagtgagagacacatta 1181
Qy 1324 caactaaaccaagaagaagatcctccctataaatttaactatgtttacaggggaat 1383
Dh 1182 caactaaaccaagaagaagatcctccctataaatttaactatgtttacaggggaat 1241
Qy 1384 gcgtacatttggtgttcccgagatttgcctcaacatgttttgaagaacccctaattccaa 1443
Dh 1242 gcgtacatttggtgttcccgagatttgcctcaacatgttttgaagaacccctaattccaa 1301
Qy 1444 caactgttgaatggttaaaagacactatagccagatgacacactctgtggccacctt 1503
Dh 1302 caactgttgaatggttaaaagacactatagccagatgacacacactctgtggccacctt 1361
Qy 1504 cagcgctgacggtgagatgctgtgtctgttcccaacaccccaagtaacgaatctcagac 1563
Dh 1362 cagcgctgacggtgagatgctgtgtctgttcccaacaccccaagtaacgaatctcagac 1421
Qy 1564 atgactctatgtccagagctgtgtcaagtgtcaggtcatgtgagggagacatcgtataaggt 1623
Dh 1422 atgactctatgtccagagctgtgtcaagtgtcaggtcatgtgagggagacatcgtataaggt 1481
Qy 1624 gctcctatgtcctcctctcgtggaatccacagcggtctactcgtttataggggtctgg 1683
Dh 1482 gctcctatgtcctcctctcgtggaatccacagcggtctactcgtttataggggtctgg 1541

Qy 1684 gacttgaattgagatgcttcaaaaacacacactgtgtgcccacaaagtlttgaccccaaggtta 1743
Dh 1542 gacttgaattgagatgcttcaaaaacacacactgtgtgcccacaaagtlttgaccccaaggtta 1601
Qy 1744 gatgataatgctctcagtgctgtagaagaataactaactgattataaagccactataggaact 1803
Dh 1602 gatgataatgctctcagtgctgtagaagaataactaactgattataaagccactataggaact 1661
Qy 1804 gaacttgaagacacactatagaagcgttgcctacctgtgtggggaagagcagatgacaacat 1863
Dh 1662 gaacttgaagacacacactatagaagcgttgcctacctgtgtggggaagagcagatgacaacat 1721
Qy 1864 gctcagaactgtctgagacagtggtggttgaggagaccagggcttgcacactcgttgcacacc 1923
Dh 1722 gctcagaactgtctgagacagtggtggttgaggagaccagggcttgcacactcgttgcacacc 1781
Qy 1924 tttagagtaagagagtgctgtattagatgtgtggtaagtagatcttctgcttgcacattg 1983
Dh 1782 tttagagtaagagagtgctgtattagatgtgtggtaagtagatcttctgcttgcacattg 1841
Qy 1984 ctgcttggtgagatgctgtgtctctcacccttaaccctagtagtctcctcactaact 2043
Dh 1842 ctgcttggtgagatgctgtgtctctcacccttaaccctagtagtctcctcactaact 1901
Qy 2044 ttctcaatagtgagaaatgagaaactgtgtgatataggagagtgagaaagagatagtgtgt 2103
Dh 1902 ttctcaatagtgagaaatgagaaactgtgtgatataggagagtgagaaagagatagtgtgt 1961
Qy 2104 agagcacttgaattcaatgaatgacgtgctgtgagcttccacttctgtgagagtgctgt 2163
Dh 1962 agagcacttgaattcaatgaatgacgtgctgtgagcttccacttctgtgagagtgctgt 2021
Qy 2164 tcccaataatccaggttgtgtagcgtgagaggaacttgaatgtaagaagaacacttccc 2223
Dh 2022 tcccaataatccaggttgtgtagcgtgagaggaacttgaatgtaagaagaagaac- ttccc 2079
Qy 2224 ttctgtactgttaacttaaaataata 2251
Dh 2080 ttctgtactgttaacttaaaataata 2107

RESULT 2
US-07-955-041-3
; Sequence 3, Application US/07955041
; Patent No. 5360733
; GENERAL INFORMATION:
; APPLICANT: BIERHUIZEN, MARTI FA
; TITLE OF INVENTION: A NOVEL BETAL-6
; TITLE OF INVENTION: N-ACTYLDIGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
; TITLE OF INVENTION: LEUKOSILIN AND A METHOD FOR CLONING PROTEINS HAVING
; TITLE OF INVENTION: ENZYMATIC ACTIVITY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,041
; FILING DATE: 19921001
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHRYN
; REGISTRATION NUMBER: 31,815


```

; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 248..314
; OTHER INFORMATION: /standard_name=
; OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-08-227-455-3

```

```

Query Match      15.9%; Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%; Pred. No. 9.4e-109;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

```

```

QY 772 aagaagcagagccttcacagacaccactcctccaccagagactgtgagac 831
DB 466 AAAAAAGGCCCTGGTGAGACCTGACGACTATATATAACATGACAGTCTTCTCT 525
QY 832 ttaagagctgaagaggttcatacagttccaccctgagcaagaaggggttccct 891
DB 526 TTCAATCAAGAGACGCAATATATTTGTAAGAACCCCTTAGTAAGAGAGGGGAGTTTCCA 585
QY 892 attgacatactatgtgtacatcagatgagaagatigaanaacttgaagaagctcgaagct 951
DB 586 ATAGCATATATCTATAGTGGTTATCAACAAGATTGAATGCTTGACAGAGCTGAGGAGGCC 645
QY 952 gtgtatgcccctcagaacatactgtgtccatgtgtgatagaagttcccaagaacttcc 1011
DB 646 ATCTATATGCTCAGAAATTTCTATTTGGCTTCATGTGACACAAATCCGAGATTTCTAT 705
QY 1012 aagaagcagctcagaagaattattcttcttcccaaatgtcctcaccagagactgtg 1071
DB 706 TTAGCTCAGATGATGGGATCGCTTCTCTTTTATGTAATGTCTTTTGAGCCAGCCGATTG 765
QY 1072 gtccgggtgtttatagctcctgtgtccaggtgtgcaagctacactcaactcatgtgaagac 1131
DB 766 GAGAGTGTGGTTATGATGCTGTGAGAGCCGGGTTGAGGCTGACCTCACTGATGATGAAGAT 825
QY 1132 ttgtccagagctcagtgccgttggaatactcctgaatacattgtggagcagacttccct 1191
DB 826 CTCTATGCAATGAGTGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 885
QY 1192 ataaagcagatcagagatggttccagagctcctcagaatgttgaaatggaggaatagatg 1251
DB 886 ATTAAACCAACTAGAAATTTGTCAGAGCTCAAGTTGTTAATGGAGAAACAACTG 945
QY 1252 gaatcagaggttacctcctaaacaaagaaccgcctggaataatcacttggagtagtg 1311
DB 946 GAACGAGAGAGATGCCATCCATAAAGAAAGAAAGTGAAGAGCGGTATGAGTGTGTT 1005
QY 1312 agagacacattacaactacaacaaagaagaagatcctcccttaatttaactatg 1371
DB 1006 AATGGAAAGCTGACAAACACAGGAGCTGCAAAATGCTTCTCCACTCGAAACACTCTC 1065
QY 1372 tttaacaggaatcgatcatctgtgtgtccagagattgttccaaatgltttgaaagac 1431
DB 1066 TTTTCTGTGGAGTGTCTTCTGTGTGTAGTAGGAGATATGTGGGTATGTACTACAGAAAT 1125
QY 1432 cctaaatcccaacaactgattgaaatggglaaagaacacttataagcccaagaatgaacactc 1491
DB 1126 GAAAAAATCCAAAAGTTGATGAGAGTGGGCAACAGACATACACACCTGATGATATCTC 1185
QY 1492 tgggcaacccctcagtgacggtgacggtgtgagtgctgtgtgttcccaacaccccaagtaac 1551
DB 1186 TGGGCGCCACCTCCAAAAGATTTCTGTAAGTCCCGGCTCACTCCCTGCACACCAATATAT 1245
QY 1552 gaacatcacaacatgacttatttgcagagctgtgtaaggtgagggtaactgagggagac 1611
DB 1246 GATCTATCTACATGACAGAGAGTTGTCAGGTGTCTCAAGTGGCAGTCTTTGAGAGGTGAT 1305
QY 1612 atcagataaggtgtcctcattgtcctcctgtctgtgaatccacagcgggctatctggtt 1671
DB 1306 GTTTCAGAGGTGCTCCTCCTACCGCCCTGCGATGAGATTCATGTGCGCTCAGTGTGATAT 1365
QY 1672 tatggggctggggagacttgaattgtagtcttcaaaaacatcaactgttggccaacagttc 1731

```

```

DB 1366 TTCGAGCTGTGACTGAGACTGATGCTCGCAAAACACCACTGTTGGCAATTAAGTTT 1425
QY 1732 gaccacaaagtagatgtaatgtctcagtgctgtcagtagaagaataaccagtttaagac 1790
DB 1426 GACGTGATGTGACCTCTTTGGCATCCAGTGTGGATGACATTTGAGACACAAAGC 1484

```

```

RESULT 4
US-08-472-482-3
; Sequence 3, Application US/08472482
; Patent No. 5658778
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, MINORU
; APPLICANT: BIERHUIZEN, MARTI FA
; TITLE OF INVENTION: A NOVEL BETA1-6
; TITLE OF INVENTION: N-ACETYLGLYCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
; TITLE OF INVENTION: LEUKOSTALIN AND A METHOD FOR CLONING PROTEINS HAVING
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,482
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/955,041
; FILING DATE: 01-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHRYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9294
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 220..1504
; FEATURE:
; NAME/KEY: polyA-signal
; LOCATION: 1913..1918
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 248..314
; OTHER INFORMATION: /standard_name=
; OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-08-472-482-3

```

```

Query Match      15.9%; Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%; Pred. No. 9.4e-109;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;
QY 772 aagaagcagagccttcacagacaccactcctccaccagagactgtgagac 831
DB 466 AAAAAAGGCCCTGGTGAGACCTGACGACTATATATAACATGACAGTCTTCTCT 525

```

```

Db 466 AAAAAAGCCCTCGGTGGACACCTGACGACTATATAACATGACCACTGACTTCTTCT 525
Oy 832 ttcaagctgaaagaagtctacagttccacgtgacgaagaaggttgagttccct 891
Db 536 TTCAATCAAGAGAGCGCAAAATATATTGTAGAAACCCCTTAGTAAAGAGAGCGGATTTTCCA 585
Oy 892 attgatactactatgtatctatcagaagattgaagaacttgaagaagcttgaagct 951
Db 586 ATAGCATATTCTATAGTGTGTTTCATCAAGATTGAATGCTTGACAGAGGTGTGAGGGCC 645
Oy 952 gtgtatgccctcagaacatatactgtgtccatgtgtgataagaagttccccaagaacttc 1011
Db 646 ATCTATATGCCCTCAAGAAATTTCTATTGCGTTGATGTGACCAAAATCCGAGATTCTCTAT 705
Oy 1012 aaagagcggtcaagaagaactatattctgtctcccaaatgtcttcataagcgtaagctg 1071
Db 706 TTAGCTGCAGTGTATGGGATCGCTTCTGTTTACTATAGTCTTGTGTGCGACGCGGATTG 765
Oy 1072 gtccggtgttattatgctcctcgtgccaggtgtcaagcttgacctcaactgtcaatgaaagac 1131
Db 766 GAGAGTGTGTATTATGCAATCGTGAGCGGGTTCAAGGCTGACCTCACTGCAATGAAGAT 825
Oy 1132 ttgtccagagctcaagtgccgtgtgaaatactcctgaatacatgtgtggaagacttccct 1191
Db 826 CTCTATGCAATGAGTGCACAACTGGAAGTACTGTATAAATCTTTGTGATGATTTTCCC 885
Oy 1192 ataagagcaatgcagaagaatgttccaggtctcagaatgttgaaatgtggaagagatgcatg 1251
Db 886 ATTTAAACCAACCTTGAATATGTCTCAGGAAAGCTCAAGTTGTATATGAGAAACAAACCTG 945
Oy 1252 agtcaaggttactcctcctaagaacaaacccgctgtgaaatcattcaacttgagttagtg 1311
Db 946 GAAAGCGAGAGAGTCCATCCCATTAAGAGAAAGTGTGAACACCGGATATAGGTCGTT 1005
Oy 1312 agagacacattacactcaacaaagaagaagatcctcccttataatlaactatg 1371
Db 1006 AATGGAAGCCTGCACAAACACAGGAGCTGTCAAAATGCTTCCCTCCACTGCAACACCTCTC 1065
Oy 1372 ttacaggaagatgtgtacatgtgtgtctcccgagattgtcccaacatgttlttgaaagac 1431
Db 1066 TTTTCTGGCAGTCTTACTCTGTGTGTCAGTATGAGGATATGTGGGTATGTACTTAACGANT 1125
Oy 1432 cctaaatcccaacaactgtattgaatgtgttlaaaagacattatagccagatgaacactc 1491
Db 1126 GAAAAAATCCAAAAGTTGATGAGTGGGCACAAACACATACAGCCCTGATAGTATCTC 1185
Oy 1492 tgggcccacccctcagcgtgcgaaggtgtgatagtcgtgtctgttcccaacccaagaagtac 1551
Db 1186 TGGGCCCACCATCCAAAGATTTCTGAAAGTCCCGGGCTCACTCCCTGCCAGCCCATAGTAT 1245
Oy 1552 gacatctcagacatgacttctatctcagagctgtgtcgaagtgtgcaggtgtcaatgaaggagac 1611
Db 1246 GATCTATCTGACATGCAAGCAGATTGCGAGTTTGTCAAGTGGCACTACTTTAGGGGTGANT 1305
Oy 1612 atcgataaggtgtcctctatgtctcgtctgttgaaatcccaacgaaggtgtatcgtgct 1671
Db 1306 GTTTCGAAGGGGTCTCCCTACCCGCGCTGCGATGAGTGCATGTCGCGCTCACTGTGCAATT 1365
Oy 1672 tatgtggtgtgaggtgtgaattgtgattgtctcaaaacacatcactgtgtgtgcaacaagttt 1731
Db 1366 TTGGAGGTGTGACTTGAATCGATGCTGCCAAACACCATTTGTTCCTAATTAAGTTT 1425
Oy 1732 gaccacaaggtgataatgtatgtctctcagtgcttgaagaatactactagttataagac 1790
Db 1426 GACGTGGAATGTTGACCTCTTTCCTCCATCCAGTGTGATGAGCAATTTAGACACAAAGC 1484

```

```

RESULT 5
US-08-487-069-3
: Sequence 3, Application us/08487069
: Patent No. 5684134
: GENERAL INFORMATION:
: APPLICANT: FUKUDA, MINORU

```

```

: APPLICANT: BIERHUIZEN, MARTI FA
: TITLE OF INVENTION: A NOVEL BETA1-6
: TITLE OF INVENTION: N-ACETYLGLUCOSAMININTRANSFERASE, ITS ACCEPTOR MOLECULE,
: TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
: TITLE OF INVENTION: ENZYMATIC ACTIVITY
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CAMPBELL AND FLORES
: STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
: CITY: SAN DIEGO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: us/08/487,069
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/955,041
: FILING DATE: 01-OCT-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: CAMPBELL, CATHRYN
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 9294
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-535-9001
: TELEFAX: 619-535-8949
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2105 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 220..1504
: FEATURE:
: NAME/KEY: POLYA_signal
: LOCATION: 1913..1918
: FEATURE:
: NAME/KEY: misc_signal
: LOCATION: 248..314
: OTHER INFORMATION: /standard_name=
: OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-08-487-069-3

```

```

Query Match 15.9%; Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%; Pred. No. 9,4e-109;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;
Oy 772 aagaagcgagagccttccacgaacacccactcctcctcaccagaagactgtgagac 831
Db 466 AAAAAAGCCCTCGGTGGACACCTGACGACTATATAACATGACCACTGACTTCTTCT 525
Oy 832 ttcaagctgaaagaagtctacagttccacgtgacgaagaaggttgagttccct 891
Db 536 TTCAATCAAGAGAGCGCAAAATATATTGTAGAAACCCCTTAGTAAAGAGAGCGGATTTTCCA 585
Oy 892 attgatactactatgtatctatcagaagattgaagaacttgaagaagcttgaagct 951
Db 586 ATAGCATATTCTATAGTGTGTTTCATCAAGATTGAATGCTTGACAGAGGTGTGAGGGCC 645
Oy 952 gtgtatgccctcagaacatatactgtgtccatgtgtgataagaagttccccaagaacttc 1011
Db 646 ATCTATATGCCCTCAAGAAATTTCTATTGCGTTGATGTGACCAAAATCCGAGATTCTCTAT 705

```

```

QY 1012 aaagagcggtcaagaataattcttgcctcccaaatgtcttcataagccagtaagctg 1071
D 706 TTAGCTGCACTGATGGGCATCGCTCTGTTTATGTAATGCTTTTGCGCCAGCGATG 765
QY 1072 gtccgggtgttatagctccctgtgccagggtgcaagctaccacactcatgaaagac 1131
D 766 GAGAGTGTGTTTATGATGCTGTGAGCCGGGTTTCAGGCTGACCTCAACTGCATGAAGAT 825
QY 1132 ttgctccagagctcagtgccgttggaataacttcctgaatacaigtggaagcagctt 1191
D 826 CTCATATGCAATGAGTGAACCTGGAAGTACTGTATTAATCTTTGTGTATGATGATTTTCC 885
QY 1192 ataaagagcaatgcagagatggtccaggctctcaagatgttgtaagaggaaatagcatg 1251
D 886 ATTAAACCAACCTAGAAATGTTTCAGAACCTCAAGTTGTTAAAGGAGAAACAAACCTG 945
QY 1252 gagtcgaaggtacctccctaagcaaaagaaaccgcgtggaatcacttgaattgagtg 1311
D 946 GAAACGAGAGATGATGCTATCCCATTAAGAAAGAGGAGAAAGCGGTATGAGGTGTT 1005
QY 1312 agagacacattacacctaaccaaaagaagaagatccctcccttaataatctaactatg 1371
D 1006 AATGGAAGAGTGACAAACACAGGAGCTGTCAAAATGCTTCTCCACACGGAACACCTCTC 1065
QY 1372 ttacagaggaatgcatcatgttggtctcccgagattcgtccaagatgtttgaagac 1431
D 1066 TTTTCTGCGAGTGCTACTTCGTGTCAGTAGGAGATATGGGGTATGTAATACAGAAAT 1125
QY 1432 cctaatacccaaacacagcttgtaagtgtaaaagacactatagccagatgaaacactc 1491
D 1126 GAAAAAATCCAAAAGTTGATGAGTGGGACAAACACATACACCCCTGTGATGATATCTC 1185
QY 1492 tgggccaacccctcagctgacagtgagatgctgctgtcttcccaacacccccaaagt 1551
D 1186 TGGGCCACATCCAAAGATATCCGGAAGTCCCGGCTCAGCTCCCTGCGACGATATATAT 1245
QY 1552 gacatctcaacatgactcttatgtccaggctgtgcaagtggcagggtcatgaggaagac 1611
D 1246 GATCTATCTACATGCAAGAGTAGTCCAGGTTTGTCAAGTGCGACTTTTGAAGGTGAT 1305
QY 1612 atcgataaggtgctcctctatgtcctgtctgtgaatcccaagcgaggatctgagct 1671
D 1306 GTTTCGAAGGCTGCTCCCTACCGCCCTGCGATGAGATGATGATGCGCTCATAGTGTGAT 1365
QY 1672 tatgggctgggagctgaattgtagtgcctcaaaacacatcactgttggccaagaattc 1731
D 1366 TTGCGACTGCTGACTGAGACTGATGCTGCGCAACACACACTGTTGCCAATAGATT 1425
QY 1732 gaccacaagtagatgataatgtctctcagtgcttagaagaatacctcagttataagac 1790
D 1426 GACGTGATGTGACCTCTTTGCCATCCAGTGTGATGAGCAATTGAGACACAAAGC 1484

```

RESULT 6
US-09-063-237-3

```

; Sequence 3, Application US/09063237
; Patent No. 6124267
; GENERAL INFORMATION:
; APPLICANT: McEver, Roger P.
; APPLICANT: Cummings, Richard P.
; TITLE OF INVENTION: O-Glycan Inhibitors of Selectin Mediated
; TITLE OF INVENTION: Inflammation Derived from P5G1-1
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center, 1201 West Peachtree
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30306-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/063,237
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/649,802
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRP110CIP7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)873-8794
; TELEFAX: (404)873-8795
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-063-237-3

```

Query Match 15.3%; Score 355; DB 3; Length 2102;

Best Local Similarity 59.3%; Pred. No. 4,2e-104; Matches 604; Conservative 0; Mismatches 415; Indels 0; Gaps 0;

```

QY 772 aagaagcgagagccttcacagacaccactacctccctccacagagaactgtgagac 831
D 465 AAAAAAGCGCCCTGGGTGACACCTGACGACTATATAAATGACAGCACTGTTCTTCT 524
QY 832 ttcaaggtgaaaggaatgcatcagttccactgagcaagaaggggtgagttccct 891
D 525 TTCACTAAGAGACCAATATATTGTAGAACCCCTTGTAAAGAGGGCGAGATTGCA 584
QY 892 attgacactcattggtgattcagagaagatgaaacacttgaaagctacagcagct 951
D 585 ATGACATATCTATAGTGTGTTCAACACAGATTAATGCTTGACAGCTGCTGAGGCC 644
QY 952 gtgtatgcccccaagaacataactgtgtcatgtgagatgagaaggtccccaagaacttc 1011
D 645 ATCATATGCTCTCAGATTTCTATTTGCTTCAATGTGACACAAATGCGAGATTCTAT 704
QY 1012 aaagagcggtcaagaataattcttgcctcccaatgtcttcacagcagtaagctg 1071
D 705 TTAGCTGCACTGATGGGCATCGCTCTGTTTATGTAATGCTTTGTGCGCAGCGCATG 764
QY 1072 gtccgggtgttatagctccctgtgccagggtgcaagctaccacactcatgaaagac 1131
D 765 GAGAGTGTGTTTATGATGCTGTGAGCCGGGTTTCAGGCTGACCTCAACTGCATGAAGAT 824
QY 1132 ttgctccagagctcagtgccgttggaataacttcctgaatacaigtggaagcagctt 1191
D 825 CTCATATGCAATGAGTGAACCTGGAAGTACTGTATTAATCTTTGTGTATGATGATTTTCC 884
QY 1192 ataaagagcaatgcagagatggtccaggctctcaagatgttgtaagaggaaatagcatg 1251
D 885 ATTAAACCAACCTAGAAATGTTTCAGAACCTCAAGTTGTTAAAGGAGAAACAAACCTG 944
QY 1252 gagtcgaaggtacctccctaagcaaaagaaaccgcgtggaatcacttgaattgagtg 1311
D 946 GAAACGAGAGATGATGCTATCCCATTAAGAAAGAGGAGAAAGCGGTATGAGGTGTT 1004
QY 1312 agagacacattacacctaaccaaaagaagaagatccctcccttaataatctaactatg 1371
D 1006 AATGGAAGAGTGACAAACACAGGAGCTGTCAAAATGCTTCTCCACACGGAACACCTCTC 1064
QY 1372 ttacagaggaatgcatcatgttggtctcccgagattcgtccaagatgtttgaagac 1431

```

```

Db 1065 TTTTCTGGACGTGCTACTTCCGTGCTCACTAGTAGGACTATGTGGGTATGTACTACAGAAAT 1124
QY 1432 cctaaatcccaactgaatgtaagaaagacactatagcccgagataaaccc 1491
Db 1125 GAAAAATCCAAAGTTGAGTAGTGCGCAAGACACATACAGCCCTGATGATGTC 1184
QY 1492 tgggacacctcagcgtgacggtgagctgctgctcccaaccaccccaagtc 1551
Db 1185 TCGCGCACATCCAAAGATTCCTGAAGTCCCGGCTCACTCCCTCCAGCATTAAGTAT 1244
QY 1552 gacatcagacatgactctatctgacagctgctgacaggtgacaggtgacagac 1611
Db 1245 GATCTATGTGATGCAACAGTGTGCGACAGTTGTCAGTGCACACTTTCAGGGTAT 1304
QY 1612 atcgataaaggtgctcctatgctcctgctcctgacatccaccagggctatcgctt 1671
Db 1305 GTTTCGAAGGGTCTCCCTACCCGCTCGATGAGTCCATGTCGCTCACTGTGCATTT 1364
QY 1672 tatgggctggggaacttgatctgacgtctcaaaacacacactgctgacacaaqttt 1731
Db 1365 TTCGGAGCTGCTGACTGTGACAGTGCCTGCGCAACACACTTGTTCCTCAATACCTT 1424
QY 1732 gacccaaggtatgataatgctcctcctcagtgcttagaagaatccctacgtataaagc 1790
Db 1425 GACGTGCATGTGACCTCTTTCCTCCATCCAGTGTTCATGACATTTGAGACAAAGC 1483

```

RESULT 7

US-08-118-906-13

Sequence 13, Application US/08118906

Patent No. 5484590

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru

TITLE OF INVENTION: Expression of the Developmental I

TITLE OF INVENTION: Antigen By a Cloned Human CDNA Encoding a Member of a

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSER: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/118,906

FILING DATE: 09-SEP-1993

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 1807 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 255..1454

US-08-118-906-13

Query Match 7.8%; Score 180.4; DB 1; Length 1807;
 Best Local Similarity 54.3%; Pred. No. 8.5e-48;
 Matches 395; Conservative 0; Mismatches 321; Indels 12; Gaps 1;

```

QY 823 tgtgacactcaaggtgtaagaagttacacagttcccaactgagcaagaagtg 882
Db 468 TCGAAGAAATTAAGTCCAGAGCCACTATACAGACGCCCTTTATGAAGAGAGAGCT 527
QY 883 gagttccctatgacactctatggtatcctatggaagattgaaacttgaaagct 942
Db 528 GACTTCCCTTGGCATATTAATGTCATCATCATCTTTCAGACACCTTTCAGAGCTC 587
QY 943 ctgagagctgtgtagccctcagacacatctctgtcactggtatgataagtcacca 1002
Db 588 TTCAGAGGCTATTTACATGCGCCCAAAATATCTCTGTGTCATGATGATGATGATG 647
QY 1003 gaactctcaagagcggtcgaagaacatctctctgctcccaatgctctcagcc 1062
Db 648 ACTGAATTTAAAGATGCGGTAGAGCAATTAAGCTGCTTCCAAAGCCTTTCTGGCT 707
QY 1063 agtaagctggtcgggtgtatgctcctcctggtccaggtgcaagctgacactcagc 1122
Db 708 TCCAGATGGAACCCGTTCTCTATGAGGAGATCTCCAGGCTCCAGGCTGACCTGAAGTGC 767
QY 1123 atggaagactgctcagagctcagtgccgtggaatcctctgataacatgtagagc 1182
Db 768 ATCAGAGATCTTCTGCTTCAGAGTCTCATGGAATACGTTATCAACACCTGTGGCA 827
QY 1183 gacttccctataagaagcaatgtagagatgctcaggtcctcaagaatgttgaatggag 1242
Db 828 GACTTCCCTCCGTAAGAAACCAAGAAATAGTTCAAGTATGTAAGAGATTTAAAGTAA 887
QY 1243 aatgcatgtagtcaaggttacctcctaaagcacaagaaccgcgtggaataat----- 1296
Db 888 AATATCCACCCAGGGGTGCTGCCCGAGCTCATGCAATTGACGAGCACTAAATATGTCAC 947
QY 1297 -----cacttggagtgtagagagacacattacacactcaacaagaagaagatcct 1350
Db 948 CAAAGACACTGGGCAAAAGCTTCTCTATGATTAAGAAACAAAGCCTTGAACCGGCT 1007
QY 1351 cccctataacttaactatgtttacaggaatgctgacatgtgctcccgagattc 1410
Db 1008 CCCCCCATTAATCTCAATTAATTAATTTAGCTTGCTGCTTATGTCCTATCAAGAGACTT 1067
QY 1411 gtccaaatggtttgagaagacccctaaatcccaacatgataaggttaagaagact 1470
Db 1068 GCAACCTTGTCTGATGACCCACGCGCTGTGATTTGCTCCAGTGCCAAAGGACACT 1127
QY 1471 tatagccagatgaacactctggccacccttcagcgtgcaagtgtagctgctgct 1530
Db 1128 TTCAGTCTGATGAGCATTTCTGAGTACATCACTAATAGATTCAGAGTCTCTGCTCT 1187
QY 1531 gtcccca 1538
Db 1188 ATGCCAAA 1195

```

RESULT 8

US-08-486-196-13

Sequence 13, Application US/08486196

Patent No. 5731420

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru

TITLE OF INVENTION: Expression of the Developmental I

TITLE OF INVENTION: Antigen By a Cloned Human CDNA Encoding a Member of a

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSER: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

US-08-486-196-13


```

Oy 823 tgtgagcacttcaagcctgaaagagatcatatcagcttccacatgagcaagaagtg 882
Db 468 TGCAAGAAATCTTGACCCAGAGCCATACATACAGACCCCTTTATCAAGAAAGAGT 527
Oy 883 gagttccctatgtcactctatctatgtgatctatgagaagatgaaacttgaaagcta 942
Db 528 GACTTTCCTTGCGATATATATATATATATATATATATATATATATATATATATAT 587
Oy 943 ctgagagctgtgtatgagccctcaagacatatatctgttccatgtgtatgagaagc 1002
Db 568 TTCAAGGCTATTTTACATGCGCCCAAAATATCTATCTGTTCATGTGATGAAAAACACA 647
Oy 1003 gaaacttcaagaagcggtgcaagaactatcttctgttccccaatgtcttcaagc 1062
Db 648 ACTGAATTTAAAGATGCGGTAGACCACTATTAAGCTGTGCTCCCAAGCGCTTTTGCGCT 707
Oy 1063 agtaagctgtgtggtgtatgtatgctccctcctggtccagggggtgcaagctgaactc 1122
Db 708 TCCAAGATGGAACCCGTTGTCTATGTAGAGGATCTCCAGGCTCCAGGCTGACCTGAAGTGC 767
Oy 1123 atggaagactgtctcagagctcagtcggtgtgaaatacttctgaatactgtgtgagc 1182
Db 768 ATCAAGATCTTTTGCTTCAGAGTCTCATGTGAAGTACGTTATACACCTGTGGGCA 827
Oy 1183 gacttccataaagaagaatgcaagatggtccaggtctcagaatgttgaatggag 1242
Db 828 GACTTCCCGCTGAACCAACAGAAATAGTTAGTATCTGAAGATTTAAAGGTAA 887
Oy 1243 aatgcatgagtgagtgactcttccctaagcaagaaccgctgtgaatat----- 1296
Db 888 AATATCACCCCAAGGGGTGCTGCCACCTCATGTGCAATTTGACGACGACTAAATATGTCCAC 947
Oy 1297 -----cacttggagtggtgagagacacatatcacttaaccacaagaagaagatc 1350
Db 948 CAGAGCACTCTGGCAAGAGCTTTCCTATGTGATGAAGAACACAGCGTGAACCGCT 1007
Oy 1351 ccccttaataatgaatgttcaaggaatgctacatgtgtgtcccgagatctc 1410
Db 1008 CCCCCCAATATCTACAAATTTACTTTGGCTGTGCTATGTGCTCTATCAAGAAAGTTT 1067
Oy 1411 gtcaaatgttgaagaacccctaataatcccaaacatgtatgaaatggtgaaagact 1470
Db 1068 GCAACTTGTGTCTCAATGACCCACGCGCTGTGTGATTTGCTCCAGTGTGCAAGACACT 1127
Oy 1471 tataagcagaatgaacactctgggcaacccctcagcgctgacggtgtatgtctgtc 1530
Db 1128 TTCAATCTCTGATGACATTTTGTGGTGAACACTCAATAGAGATTCCAGTGTCTGTGCT 1187
Oy 1531 gtcccaa 1538
Db 1188 ATGCCAA 1195

```

RESULT 10
US-08-474-065-13
Sequence 13, Application US/08474065

```

: GENERAL INFORMATION:
: APPLICANT: Fukuda, Minoru
: APPLICANT: Bierhuizen, Marti F.A.
: TITLE OF INVENTION: Expression of the Developmental I
: TITLE OF INVENTION: Expression By a Cloned Human cDNA Encoding a Member of a
: TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell and Flores
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

```

: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/474,065
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/118,906
: FILING DATE: 09-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 9526
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1807 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 255..1454
: US-08-474-065-13

```

```

Query Match 7.8%; Score 180.4; DB 2; Length 1807;
Best Local Similarity 54.3%; Pred. No. 8.5e-48;
Matches 395; Conservative 0; Mismatches 321; Indels 12; Gaps 1;
Oy 823 tgtgagcacttcaagcctgaaagagatcatatcagcttccacatgagcaagaagtg 882
Db 468 TGCAAGAAATCTTGACCCAGAGCCATACATACAGACCCCTTTATCAAGAAAGAGT 527
Oy 883 gagttccctatgtcactctatctatgtgatctatgagaagatgaaacttgaaagcta 942
Db 528 GACTTTCCTTGCGATATATATATATATATATATATATATATATATATATATATAT 587
Oy 943 ctgagagctgtgtatgagccctcaagacatatatctgttccatgtgtatgagaagc 1002
Db 568 TTCAAGGCTATTTTACATGCGCCCAAAATATCTATCTGTTCATGTGATGAAAAACACA 647
Oy 1003 gaaacttcaagaagcggtgcaagaactatcttctgttccccaatgtcttcaagc 1062
Db 648 ACTGAATTTAAAGATGCGGTAGACCACTATTAAGCTGTGCTCCAAAGCGCTTTTGCGCT 707
Oy 1063 agtaagctgtgtggtgtatgtatgctccctcctggtccagggggtgcaagctgaactc 1122
Db 708 TCCAAGATGGAACCCGTTGTCTATGTAGAGGATCTCCAGGCTCCAGGCTGACCTGAAGTGC 767
Oy 1123 atggaagactgtctcagagctcagtcggtgtgaaatacttctgaatactgtgtgagc 1182
Db 768 ATCAAGATCTTTTGCTTCAGAGTCTCATGTGAAGTACGTTATACACCTGTGGGCA 827
Oy 1183 gacttccataaagaagaatgcaagatggtccaggtctcagaatgttgaatggag 1242
Db 828 GACTTCCCGCTGAACCAACAGAAATAGTTAGTATCTGAAGATTTAAAGGTAA 887
Oy 1243 aatgcatgagtgagtgactcttccctaagcaagaaccgctgtgaatat----- 1296
Db 888 AATATCACCCCAAGGGGTGCTGCCACCTCATGTGCAATTTGACGACGACTAAATATGTCCAC 947
Oy 1297 -----cacttggagtggtgagagacacatatcacttaaccacaagaagaagatc 1350
Db 948 CAGAGCACTCTGGCAAGAGCTTTCCTATGTGATGAAGAACACAGCGTGAACCGCT 1007
Oy 1351 ccccttaataatgaatgttcaaggaatgctacatgtgtgtcccgagatctc 1410
Db 1008 CCCCCCAATATCTACAAATTTACTTTGGCTGTGCTATGTGCTCTATCAAGAAAGATT 1067

```


Matches 239; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```

Oy 862 ccactgagaagaaggtgagttccctattgcatactcatgtgataatgagaag 921
    || || || || || || || || || || || || || || || || || || ||
Db 13 CCCCTTAGTAAGAGAGAGCGGAGTTTCCAAATGACATATTCTATGTTGTTCAAG 72
Oy 922 attgaaacttgaagactgagctgtgtatgccccccaacatactatctgac 981
    || || || || || || || || || || || || || || || || || || ||
Db 73 ATTGAATGCTTGACAGGCTGCTGAGAGGCCATCATATATGCTCAGAAATTTCTATTCGTT 132
Oy 982 catgtgagatgagatgccccaagaacttcaagaaggcggtcaagaactatcttcgc 1041
    || || || || || || || || || || || || || || || || || || ||
Db 133 CATGTGGACCAAAATCCAGAGATTCCTATTAGCTGATGAGGATCCGCTTCCTGT 192
Oy 1042 tccccaaagtcttcaagcaagtaagctggttcggtgtatgccccctggtccag 1101
    || || || || || || || || || || || || || || || || || || ||
Db 193 TTTAGTAATGCTTTGTGGCCAGCCGATTCGAGAGTGTGTTATGATCAGTGAGACCGG 252
Oy 1102 gtgcaagctgacctcaactgcatggaagactgtctccagactcagtcggtggaatac 1161
    || || || || || || || || || || || || || || || || || || ||
Db 253 GTTCAGGCTGACCTCACTGACATGAGAGATCTCTATGCAATGATGCAAACTGGAAGTAC 312
Oy 1162 ttccctgaatacatgtggaagcgaacttctctataaagaagaatgcaagatgtgccagct 1221
    || || || || || || || || || || || || || || || || || || ||
Db 313 TTGATTAATCTTTGTGTGATGATTTTCCCATTAATAACCACTAGAAATGTGCAGGAG 372
Oy 1222 ctcaag 1227
    || || || || || || || || || || || || || || || || || || ||
Db 373 CTCGAG 378

```

RESULT 13

US-08-488-135-3
; Sequence 3, Application US/08488135
; Patent No. 5786910

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru
APPLICANT: Bierhuizen, Marti F.A.
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,135
FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
type: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-488-135-3

Query Match 7.0%; Score 162.8; DB 1: Length 378;
Best Local Similarity 65.3%; Pred. No. 1,4e-42;
Matches 239; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```

Oy 862 ccactgagaagaaggtgagttccctattgcatactcatgtgataatgagaag 921
    || || || || || || || || || || || || || || || || || || ||
Db 13 CCCCTTAGTAAGAGAGAGCGGAGTTTCCAAATGACATATTCTATGTTGTTCAAG 72
Oy 922 attgaaacttgaagactgagctgtgtatgccccccaacatactatctgac 981
    || || || || || || || || || || || || || || || || || || ||
Db 73 ATTGAATGCTTGACAGGCTGCTGAGAGGCCATCATATATGCTCAGAAATTTCTATTCGTT 132
Oy 982 catgtgagatgagatgccccaagaacttcaagaaggcggtcaagaactatcttcgc 1041
    || || || || || || || || || || || || || || || || || || ||
Db 133 CATGTGGACCAAAATCCAGAGATTCCTATTAGCTGATGAGGATCCGCTTCCTGT 192
Oy 1042 tccccaaagtcttcaagcaagtaagctggttcggtgtatgccccctggtccag 1101
    || || || || || || || || || || || || || || || || || || ||
Db 193 TTTAGTAATGCTTTGTGGCCAGCCGATTCGAGAGTGTGTTATGATCAGTGAGACCGG 252
Oy 1102 gtgcaagctgacctcaactgcatggaagactgtctccagactcagtcggtggaatac 1161
    || || || || || || || || || || || || || || || || || || ||
Db 253 GTTCAGGCTGACCTCACTGACATGAGAGATCTCTATGCAATGATGCAAACTGGAAGTAC 312
Oy 1162 ttccctgaatacatgtggaagcgaacttctctataaagaagaatgcaagatgtgccagct 1221
    || || || || || || || || || || || || || || || || || || ||
Db 313 TTGATTAATCTTTGTGTGATGATTTTCCCATTAATAACCACTAGAAATGTGCAGGAG 372
Oy 1222 ctcaag 1227
    || || || || || || || || || || || || || || || || || || ||
Db 373 CTCGAG 378

```

RESULT 14

US-08-474-065-3
; Sequence 3, Application US/08474065
; Patent No. 5830465

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru
APPLICANT: Bierhuizen, Marti F.A.
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,065
FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

```
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-474-065-3
```

```
Query Match 7.08; Score 162.8; DB 2; Length 378;
Best Local Similarity 65.3%; Pred. No. 1.4e-42;
Matches 239; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
```

```
QY 862 ccactgagcaagaagagtgaggttcctcattgcatcctatgagtgatcagagaag 921
DB 13 CCCCTTGTATAGAGAGAGCGGAGTTTCCATACATATTCTATAGTGTTCATCACAG 72
QY 922 atgaaacttgaaagagctactgagagctglatgccccctcagaacatatactgtgc 981
DB 73 ATGGAATGCTTGACAGGCTCTGAGGGCCATCTATATGCTCAGAAATTTCTATGGCTT 132
QY 982 catgtgatgagaagtcgccagaacattcaagaagcggtcaagaacattattcttgc 1041
DB 133 CATGTGACACAAATCCGAGAGTTCCTATTACCTGCAGATGGGATGCTTCCTGT 192
QY 1042 tcccaaatgctcattagcagtaagctggttcggtgttatactcctcctcagga 1101
DB 193 TTTGATATGCTTTGTTGGCCAGCTTGAGAGTGTGTTATGCACTCGTGAGCCGG 252
QY 1102 gtgaagctgacctcaactgcatgaaagactgtccagagctcagtcagtcgtggaatac 1161
DB 253 GTTCAGGCTGACCTCACTGATGAGAGATCTATGCAATGAGTGCAGAACTGGAAGTAC 312
QY 1162 tctcgtatacatgtggaaggaacttcctataaagagcaatgcagagatgtccaggt 1221
DB 313 TTGATTAATCTTTGTGTATGATTTCCATTAAACCACTGAATATGTCAGGAG 372
QY 1222 ctcaag 1227
DB 373 CTCAG 378
```

```
RESULT 15
US-08-118-906-1
; Sequence 1, Application US/08118906
; Patent No. 5484590
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Bierhuizen, Marti F.A.
; TITLE OF INVENTION: Expression of the Developmental I
; TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
; TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; City: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,906
FILING DATE: 09-SEP-1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-118-906-1
```

```
Query Match 5.88; Score 134; DB 1; Length 378;
Best Local Similarity 60.4%; Pred. No. 2.8e-33;
Matches 221; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
```

```
QY 861 cccactgagcaagaagagtgaggttcctcattgcatcctatgagtgatcagagaag 920
DB 12 CCCCTTGTATAGAGAGAGCGGAGTTTCCATACATATTCTATAGTGTTCATCACAG 71
QY 921 gattgaaacttgaaagagctactgagagctglatgccccctcagaacatatactgtc 980
DB 72 CTTTGACACCTTTGCCAAGGCTCTTCAGGCTATTACATGCCCCAATATCTCTGTGT 131
QY 981 ccatgtgatgagaagtcgccagaacattcaagaagcggtcaagaacattattcttgc 1040
DB 132 TCATGTGATGAAAGAAACCAACACTGAATTTAAAGATGCGGTAGACACTATTAAGCTG 191
QY 1041 cttccaaagtctcattagcagtaagctggttcggtgttatactcctcctcagga 1100
DB 192 CTTCACCAAGCCTTTCTGCTTCCAAAGATGGAACCCGTTCTATGAGAGGATCTCAG 251
QY 1101 ggtgaagctgacctcaactgcatgaaagactgtccagagctcagtcagtcgtggaata 1160
DB 252 GCTCCAGGCTACCTGAACATCATCAGAGATCTTCTGCTTCGAGGTCTCATGGAAGTA 311
QY 1161 tctcgtatacatgtggaaggaacttcctataaagagcaatgcagagatgtccaggt 1220
DB 312 CGTTATCAACACCTGTGGCAAGACTTCCCTGAAACCAAGGAATATGTTACAGTA 371
QY 1221 tctcaa 1226
DB 372 TCTGAA 377
```

```
Search completed: September 26, 2002, 08:12:36
Job time: 13541 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 03:04:05 : Search time 2719.46 Seconds
(without alignments)
11509.432 Million cell updates/sec

Title: US-09-874-390-1
Perfect score: 2319
Sequence: 1 atactaggttctctatt.....ataataaacttaacaga 2319

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estbta:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	923.6	39.8	1862	11 AK008234	Mus muscu
2	812.2	35.0	864	10 BG821420	BG821420
3	809.8	34.9	826	10 BG821370	BG821370
4	757	32.6	872	10 BG747491	BG747491
5	739.8	31.9	1003	10 BG385575	BG385575
6	711	30.7	875	10 BG470202	BG470202
7	704.8	30.4	766	10 BG386598	BG386598
8	693.4	29.9	866	10 BG468641	BG468641
9	691.6	29.8	731	10 BG248304	BG248304
10	685.6	29.6	702	10 BG386247	BG386247
11	666.2	28.7	920	10 BG248304	BG248304
12	653	28.2	871	10 BG386293	BG386293
13	651.8	28.1	736	10 BE617113	BE617113
14	650.8	28.1	955	10 BG328280	BG328280
15	646.4	27.9	671	10 BG750604	BG750604
16	640	27.6	640	9 AA583339	AA583339
17	640	27.6	640	9 AA583339	AA583339

18	633.4	27.3	842	10 BE869192	BE869192
19	622.6	26.8	655	9 AM842622	AM842622
20	615.6	26.5	882	10 BG171085	BG171085
21	613.6	26.5	850	10 BG822181	BG822181
22	613.2	26.4	724	10 BG469448	BG469448
23	611	26.3	612	10 BE292814	BE292814
24	608.4	26.2	912	10 BE871910	BE871910
25	607.2	26.2	1050	10 BE867668	BE867668
26	605	26.1	671	9 A1587061	A1587061
27	599.8	25.9	603	9 AM192901	AM192901
28	597.4	25.8	978	10 BG171515	BG171515
29	586.8	25.3	629	9 BG167918	BG167918
30	582.6	25.1	606	9 AM662377	AM662377
31	557.2	24.0	823	10 BG169857	BG169857
32	556.6	24.0	763	10 B1826575	B1826575
33	552	23.8	659	10 BE304708	BE304708
34	539	23.2	583	9 AM964294	AM964294
35	516.4	22.3	554	9 AA307800	AA307800
36	480.2	20.7	609	10 BE881417	BE881417
37	477.2	20.6	904	10 BG469649	BG469649
38	436	18.8	436	9 AA568218	AA568218
39	433	18.7	433	9 AA583146	AA583146
40	426.4	18.4	433	9 AA315469	AA315469
41	406	17.5	830	10 BE868512	BE868512
42	383	16.5	383	9 A1150400	A1150400
43	381.2	16.4	567	12 A2388491	A2388491
44	377.4	16.3	420	9 AM374998	AM374998
45	375.2	16.2	965	10 BF582709	BF582709

ALIGNMENTS

RESULT 1	AK008234	1862 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	AK008234				
DEFINITION	Mus musculus adult male small intestine cDNA, RIKEN full-length				
	enriched library, clone:2010013H22:homolog to				
	BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE, full insert sequence.				
ACCESSION	AK008234	GI:12842295			
VERSION	AK008234.1				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library				
	clone:2010013H22.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2 (sites)				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				
REFERENCE	3 (sites)				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishii, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiki, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multipipillary sequencer				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				

[illegible]

RESULT	3
BG821370	
LOCUS	BG821370
DEFINITION	602724914r1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE4864634 5', mRNA sequence.
ACCESSION	BG821370
VERSION	BG821370.1 GI:14168957
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 826)		NIH-MGC http://mgc.nhl.nih.gov/ .	National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished (1999)	Contact: Robert Strausberg, Ph.D.	

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLNL at: <http://image.llnl.gov>
 Plate: LICM1728 row: k column: 03
 High quality sequence stop:823.

FEATURES	Location/Qualifiers
source	1. .836

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4864634"
/clone_1ib="NH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCRB1; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAC(G). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

```

Query Match	34.9%	Score 809.8;	DB 10;	Length 826;
Best Local Similarity	99.6%	Pred. NO. 1.6e-224;		
Matches 822; Conservative	0;	Mismatches 2;	Indels 1;	Gaps 1;

Db	2	CACGGGAACCTGCCCCCTTGCTACTTGTGACCTGCGCCCTTATCTAGCAGTTTGTTCGGGA	61
QY	132	agccctgggattctbctlaatacctacatcactgtaagtgctgaagggaacacagatgaagac	191
Db	62	AGCCCTGGGATTCTGCTATACCTATCACTGAGTGGCTGAAGGGAAMCAGATGAAGAAC	121
QY	192	atgacctaaaggagctccctctgtaactgaagaagccaagctgaagcctggcaagaatla	251
Db	122	ATGACCTCAAGAGGCTGCTGCTGCAATGGAAGACCAACCTGACGCGCTGGCAAGATATTA	181
QY	252	aagaaagacccgaacatcttcccttggaacatcttaagaatgcagaatatactctttag	311
Db	182	AAGAGAGACCTCAAACTGTCTCCCTTGACATCTTATGAAGATGACAAATATACCTTTGGAG	241
QY	312	ggttgaagaatcaaggagacatggttgcttcaacattgtgcgcacaggaacacgcagctt	371
Db	242	GGTTAGAAGATCAGGGGACATGTTGTTTCACATTGGCTGCCACGSAACACCGCAGCTTT	301
QY	372	cacttgaagaacagaatcaagccttgtaagaagatcaatccctaagaagagaagaact	431
Db	302	CACCTTGGAAACAGATATCAGCGCTTGTGAAGAGATCATCCCTAAGCAGGAGAAAGCTACT	361
QY	432	aaagatattgtctccctcccaaccttcccttgctgctgagctcacacctgctcccaattcgt	491
Db	362	AAAGGATTGTGCTCCCTCCACCTCCCTGTCGTCGTCACCTGTCCTCCATTCTGT	421
QY	492	gaagatggttcaatgaagaagacatctgcacagtcatctactgttgagctctggagctga	551
Db	422	GACGATGGTTCAATGGAAGAGACTCTGCCAGCTGCATTACTTGTGGGCTCTGGGCTGCTA	481
QY	552	tatgtctgctgacacatgtgctctgaacattcttcaagtgtaagtgtaacctgacca	611
Db	482	TATGCTGCTGGCCACTGAGGCTCTGAACCTTCTTTCAGGTTGAAGTGTGACTGCACCA	541
QY	612	cttgagctctgagctccagggaatctc-aaagccaactgttaagaaatatctgtataat	670
Db	542	CTTGGGCTGTGGAGTCCAGGGAATTCCAAAAGCCATACGTGAAGAAATCTTGTAATAAT	601
QY	671	tccctaaactccagaagaagagctatcaacttctcaagggttaacccgagagggaacag	730
Db	602	TCTCTAAACTTCACACAAAGAAGCTATACATGTTCAAGGGCTCACCGAGGGGACCAAG	661
QY	731	aggcagtgcttcaagctatctgaataaaccttgagagttcaagaagaagcagagaccttca	790
Db	662	AGCGAGTGGCTTAGGCTATTCTGATTAACCTGAGAGTCAAGAAGAAGGAGAGCTTTCA	721
QY	791	cagacacccaactactcttccctcacagagacgttgagcaacttcaaggctgaagaagt	850
Db	722	CAGACAAACCACTACCTCTCCCTCAACAGAGACTGTGACACTCTCAAGGCTGAAGAAGAGT	781
QY	851	tcatctagttctccactttaggaagaagaagggtgagtttccattag	895
Db	782	TCATCAGTCTCCACTGAGCAAGAAGAAGGTGAGATTCCTATTG	826

RESULT	4				
LOCUS	BC747491				
DEFINITION	BC747491	872 bp	mRNA	linear	EST 15-MAY-2001
ACCESSION	60270406061				
VERSION	BC747491				
KEYWORDS	BC747491.1	GI:14058144			
SOURCE	EST.				
ORGANISM	human.				
	Homo sapiens				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 872)	NIH-MGC	http://mgc.nci.nih.gov/ .	National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished (1999)		Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: L10C1711 row: k column: 01
 High quality sequence stop: 747.
 Location/Qualifiers

FEATURES

source

1. 872
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4645159"
 /clone_lib="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the Laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 237 a 202 c 214 g 219 t
 ORIGIN

Query Match

Best Local Similarity 96.0%; Pred. No. 4,1e-209; Length 872;
 Matches 819; Conservative 0; Mismatches 30; Indels 4; Gaps 4;

QY 185 gaagaacatgacctgaagagcttcctgcaatgagaagaccagctgagcctggcaaa 244
 |||||||
 DB 1 GAAGAACAATGACCTCAAGAGCTTCCTGCATGAGAACCAAGCTGACGCTGGCANA 60
 QY 245 gattataagagagcctgaactgttctcttgacatcttcatgattgtagaataact 304
 |||||||
 DB 61 GATATTAAAGAGACCTGAACCTGTCTTGCATCTTATGAATGTCAGAAATATAC- 119
 QY 305 ttctgaggttgaagatcaaggagacatgtgttcaacattgtctgcacaggaacacgc 364
 |||||||
 DB 120 TTTCGAGGTTGAAAGATACGGGACATGTTGTTTCACATTTGCTGCCACGAAACACCC 179
 QY 365 cagcttcaacttgaaacagaatcaacgcctgtgaaagatcatccctaaagcaggaga 424
 |||||||
 DB 180 CAGTCTTACTTGGAACAGATACAGCCTTGTGAAGAGATATCCCTAAGCAGAGAGA 239
 QY 425 agctactaaagatgtgtcctcctccacattcctgtgctggtctccactgtctccc 484
 |||||||
 DB 240 AACTACTAAAGATGTGCTCCTCCACCTTCCCTGTGCTCGCTCCACCTGTCTCC 299
 QY 485 attctgtgacgtgttcaatgagaagacatctgcacgtgattactgttggtctcgg 544
 |||||||
 DB 300 ATTCTGTGACGATGCTTCATGGAAGAGACTCTGCCAGCTGATTAATCTTG6GCTTG 359
 QY 545 gctgctataatgtctgacgtcctgtgctctgaaacttcttcaagttgaagtgtgact 604
 |||||||
 DB 360 GCTGCTATATGCTGCTGGCAGCTGTGGCTCTGAACCTTCTTCAAGGTTGAAGTGA 419
 QY 605 ctgacacactgtgtgaggtccaggaatctcaaaagcagctactgttagaatactgt 664
 |||||||
 DB 420 CTGACACTTGGGCTGAGGAGTCCA-GGAATCTCAAGCAGATCTAGGAATATCTTGT 478
 QY 665 ataattctgaaacttcaagcaagaggttatcaactgttcaaggggtcaccgcaagg 724
 |||||||
 DB 479 ATATATTTCTGAAACTTCACAGCAAGAGGTCTATCACTGTTCAAGGGGTCACCCG 538
 QY 725 accaagagcagtgcttcaagctatcttgaataacctggaaggttcaagaaagcagagc 784
 |||||||
 DB 539 ACCAAGAGGCACTGCTTCAAGGCTATTTCAATTAACCTGAGAGTCAAGAGAGCGAGAC 598

QY 785 ctccacagacaccactactctccctcaccagagactgttagacttcaaggtcga 844
 |||||||
 DB 599 CTTTCACAGACACCCACAT-CTCTCCCTACACAGAGACTGTAGCACTTCAAGGCTGAA 657
 QY 845 ggaagttcaatcagttccactgagcaagaagaggtgtgagttccctattgactccta 904
 |||||||
 DB 658 GGAAGTCTATACAGTTCACAGTCCAGTCAAGAGAGAGGTGAGTTCCTATTCATCTA 717
 QY 905 tggatttcaatgagaagattgaaacttgaagagctactgtgagctgtgtat-gccct 963
 |||||||
 DB 718 TGGTATTCATGAGAGATGAAGACTTGAAGGCTCTGAGAGCTGTGATGGCCCT 777
 QY 964 cagaacataatgtgtcatgtgatgagaagctcccaagaaacttcaagagcgctc 1023
 |||||||
 DB 778 CAGAAATATTAATCTGGCTCTGTGAGTTGAGAGTTCCTCCCAAAATTTCAAAAGGGGGC 837

QY 1024 aaagaactattt 1036
 || || || ||
 DB 838 AAGCAGATATTTT 850

RESULT 5
 BG469141
 LOCUS 602511125P1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4645159 5',
 DEFINITION mRNA sequence.
 ACCESSION BG469141
 VERSION BG469141.1 GI:13401416
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 912)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: L10C1418 row: b column: 08
 High quality sequence stop: 705.
 Location/Qualifiers

FEATURES

source

1. 912
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4645159"
 /clone_lib="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the Laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 240 a 214 c 231 g 227 t
 ORIGIN

Query Match 32.6%; Score 757; DB 10; Length 912;
 Best Local Similarity 94.2%; Pred. No. 4.2e-209;
 Matches 841; Conservative 0; Mismatches 45; Indels 7; Gaps 5;

QY 74 cggagactgacctgtactgtgacgtgaccttaactcagcagtttctgttggag 133
DB 2 CGGGAACTGCGCTTGTACTCTGTGACCTGCGCTTTACTCAGCACTTTTGTTCGGAGAG 61
QY 134 cccctggattctgtactatcactatcactgtatgagtgctggaagggaaagaagaacat 193
DB 62 CCGTGGATTTGCTTAATACCTATACCTGATGAGTGTGAGGAAACAGATGAAGAACAT 121
QY 194 gacctcaaggagctctctgtcaatgagaagaacaaagctgaagccttggcaaatataa 253
DB 122 GACCTCAAGAGAGCTTCTGTCATGAGAGACCAAGCTGAGCCTGGCAAGATATTA 181
QY 254 gaggagctgaactgttctctgtgacatctaataatgacgaataacttttggagg 313
DB 182 GAGAGAGCTGAACCTGTTCTTGGACATCTTATCAATGTCGAAATAACCTTTTGGAGGG 241
QY 314 ttagaagatcagagagacatgtgtgttcaacatgtgtgccaagagaacccgacgttca 373
DB 242 TTAAAGATCAGGGGACATGTTGTTTCAATTTGCTGCCAGGAAACCGGACGCTTCA 301
QY 374 ctggaaacagaatcaagccttgtgaaagatacaccctaagcagaagaagactactaa 433
DB 302 CTGGAACAGCAATCAGCGCTTGTGAAGAGATCACCCTAAGCAGAGAGAGCTACTAA 361
QY 434 aggatgtgtctctctcactcacttccctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 493
DB 362 AGGATTTGTCTCTCTCCACCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 421
QY 494 cgaatgttcaatgagaagagactctgccaagctgacatctgtgtgtgtgtgtgtgtgt 553
DB 422 CGATGCTTCAATGAGAGAGACTCTGCCAGCTGCACTTGTGGGCTTGGGCTGCTGTA 481
QY 554 tgcgtgtgccaactgt 613
DB 482 TGCTGCTGCGCAGCTGTGCTGTGAAGCTTCTTCAAGGTGAGGTGAGCTGACCACT 541
QY 614 tggctgtgagatcagaagaatcacaagccagctactgtgagaataatctgtataatttc 673
DB 542 TGGGTCTGGAATCCA-GGAATCTCAAGCCAGTACTGAGGAATATCTTGTATATTTCC 600
QY 674 tgaacttccagcaagaagagctcactcaactgttcaaggggtcaccgagaggaagag 733
DB 601 TGAACCTTCCAGCAAGAGGCTATCAACTTTCAGGGGTACCCGAGGAGCAACAAGG 660
QY 734 cagtgtctcaagctatcttgaataaactgtgaggtcgaagaagaagcgaagccttcaag 793
DB 661 CAGTGTCTCAAGCTATCTGATTAACCTGAGGTGAGAAAGCAGAGCCTTTTTCACAG 720
QY 794 acaccctactctcctcactcaccagaga--ctgtgagcacttcaagctggaagaagtt 851
DB 721 ACACCAATAGCTTACCCCTCACAGAGAACTGTGAGCACTTCAAGGCTGAAAGGAGCT 780
QY 852 c-atacagttccactagcaagaagaaggtgtgagttccctatgcatcactatgtgt- 909
DB 781 CAATACAGTTCCACTGAGC-AAGAAGAGGTGAGGTTCCCTATAGCATACTCTATGGGG 839
QY 910 -atcatgagaagaattgaaaacttgaagaagctactgtgagctgtgtgtgtgtgtgtgt 961
DB 840 AACCATGAGAGCATGTGAACCTTGAAAGCTATGCAAGCTGTGAAGACC 892
RESULT 6
LOCUS BG385575 1003 bp mRNA linear EST 12-MAR-2001
DEFINITION 602453872F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4582504 5',
mRNA sequence.
ACCESSION BG385575
VERSION BG385575.1 GI:13278389
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1003)
AUTHORS NIH-National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCML304 row: O column: 17
High quality sequence stop: 694.
location/Qualifiers
1. 1003
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4582504"
/clone_1ib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon. Vector: pORF7. Site_1: XhoI. Site_2: EcoRI. cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(g). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 282 a 248 c 264 g 209 t
ORIGIN
Query Match 31.9%; Score 739.8; DB 10; Length 1003;
Best Local Similarity 93.1%; Pred. No. 4.6e-204;
Matches 818; Conservative 0; Mismatches 57; Indels 4; Gaps 4;
QY 174 gggaaacagatgaagaacatgacctcaagagcttccgtgcaatgagaagaacagctga 233
DB 1 GGGAAACAGATGAGAACATGACTCAAGAGAGCTTCTGTCATGAGAACACCAAGCTGA 60
QY 234 cgctgtgcaagaatataaagagagcctgaactgttctctgtgacatctatgaatgtc 293
DB 61 CGCTGTGCAAGATATTAAGAGAGAGCGCTGAACCTGCTTGCATCTTATGAATGTC 120
QY 294 agaaatacacttctggaagggttagaagatcagaaggagacatgtgtgtacattgtgcga 353
DB 121 AGAAATTAACCTTTGGAGGTTAGAAATCAAGGGAATGTTTACATTGCTGCCA 180
QY 354 cggacacccgcaagcttctcacttgaagaagaatcaagccttgtgaagaagatcacccta 413
DB 181 CGGAACACCGCAATCTTCACTTGAAGAAAGAAATCACAGCCTTGTGAAGAGATATCCCTA 240
QY 414 agcagagagaagctactaaaggattgtgtcctcctccacttccctgtgtgtgtgtgtgt 473
DB 241 AGCAGGAGAGAAGCTACTAAAGATGTTGTCCTCTCCACTTCCCTGTCGTGTCCTCC 300
QY 474 accgtgtccattctgtgacgagatggttcaatgagaagaagcctgccaagctgatatctt 533
DB 301 ACCTGTCTCCATTTCTGTGAGATGTTCAATGAGAAAGAGCTGCGCAGCTGATTACTT 360
QY 534 gtggagcttggagctgcatatgtcgtcggcgaactgtgtgctgaaacttcttcaagtt 593
DB 361 GTGGGCTCTGGGCTGCTATATGCTGCTGCGCACCTGTGCTGTAACCTTCTTTCAGGTT 420
QY 594 gaagtgtactctgacacttgggtgtgagtcgaaggaatctcaaaagcagactgtag 653
DB 421 GAACTGTGACTCTGACCACTTGGGCTGAGTCCAGGGAATTCACAAGCAGATACGTAG 480
QY 654 gaatactgtataatttcttgaacttccagcaagaaggtctataactgttccaggggt 713

Db 481 GAATATCTGTATATATTTCTGAAACTTCCAGCAAGAGGTCTATCAACTGTTCAGGGGT 540
 QY 714 caccgaaggggacccaagagagcagtcctcagcgtatcttgataaaccctggaggtcaaga 773
 Db 541 CACCCGAGGGGACCAAGAGGCACTGCTTCAGGCTATCTGATTAACCTGGAGGTCAAGAA 600
 QY 774 gaagcgagagccttcacaagacacccactacccctccctccacagagagactgtgaacatt 833
 Db 601 GAAGGAGAGGCTTTCACAGACACCCACTTACTCTCCCTCCAGACAGACTGGAGACTT 660
 QY 834 caagcgtaagaggaattcatacagttcccaactgagcaagaagaggtgagttcccat 893
 Db 661 CAAGGCTAAAGAGAGTTCATACAGTT-CCACTGAGCAGACAGAGGTGAGACTTCCCTAT 719
 QY 894 tgcac-actcattggttgattatgagaa-gattgaaacttgaaagctactggaact 951
 Db 720 TGCATAACTCTATGGGATTCACGGAAGGCTTGACACTTTCAGAGCTTAATGGGACT 779
 QY 952 gtgtatgcccctcaag-aacataactgtgtccatgtgagatgagaaagtcoccaagaactt 1010
 Db 780 GGGAGTGACCTCAACAAACATATAACGGGGCCGGTGGCAGTAAGAACCCCGAAATTC 839
 QY 1011 caaagagcggtccaagaactattcttctgtcccaaa 1049
 Db 840 ACAGAGGGGGCAACCAACATATCTGAGGTCCCAA 878

 RESULT 7
 Bg470202 875 bp mRNA linear EST 21-MAR-2001
 LOCUS 60253711F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4661355 5',
 DEFINITION mRNA sequence.
 ACCESSION Bg470202
 VERSION Bg470202.1 GI:13402477
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 875)
 NIH-MGC http://mgc.ncl.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rsb@bhs.jrmail.nih.gov
 Tissue Procurement: AMCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LNCX1460 row: e column: 04
 High quality sequence stop: 814.
 Location/Qualifiers
 1..875
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4661355"
 /clone_id="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pORF7; Site:1: XhoI; Site:2:
 EcoRI; CDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-CDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 223 a 211 c 219 g 222 t
 ORIGIN

Query Match 30.7%; Score 711; DB 10; Length 875;
 Best Local Similarity 93.5%; Pred. No. 1,1e-195;
 Matches 811; Conservative 0; Mismatches 40; Indels 16; Gaps 6;

 QY 391 gcccttgaagagatcaccctaaagcagagagaagactactaaagattgtgtcctc 450
 Db 2 gcccttgaagagatcaccctaaagcagagagaagactactaaagattgtgtcctc 61
 QY 451 caccctccctgtgtcgtgtccaccctgtcccaattctgtgaagaagtgttcaatgaag 510
 Db 62 CACCTTCCCTGTGCTGGCTCCACCTGCTCCCTGATGACGATGCTCAAGGAAG 121
 QY 511 agactctgcagctgactactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 570
 Db 122 AGACTTGCACAGCTGACTTACTTGTGGCTCTGGGCTGCTATATGCTGCTGGCCTGTG 181
 QY 571 gctctgaactctcttcaagttgaggtgtgactgtgacccactgtgtgtgtgtgtgtgt 630
 Db 182 GCTTGAAACTTTCTTTGAGGTGAAGTGTGACTGTGACCACTTGGTGGAGTCCAGG 241
 QY 631 gaattcaagcagactgtgagaaatctgtataatttccctgaactccagaag 690
 Db 242 GAATCTCAAGCAGACTGATGTAAGATATCTGTATATTTCTGAAACTTCCACCAAG 301
 QY 691 aggtctcaactgttcaaggtgtaccccgagagagcagaagagcagtgcttcaagctatt 750
 Db 302 AGGTTATCAACTGTTCAAGGGGTACCCGAGGGGACCAAGAGGCAAGTCTTCAGGCTATT 361
 QY 751 ctgaataacctgtgaagtcagaagaagcgaagccttccacagacacccactactctc 810
 Db 362 CTGAATTAACCTGAGAGTCAAGAGAGAGGAGCCTTTCACAGACACCACACTGCTGC 421
 QY 811 ctaccagagactgtgagcaacttcaaggtcgaaggaattcatacagattcccaatgagc 870
 Db 422 CTCACAGAGAGCTGTGAGACACTTCAAGGCTGAAAGAGTTCTATCAATTCCTCCACTGAGC 481
 QY 871 aaagaaggtgtgaggttccctcatctgcatctctatgtgattcatgagagaattgaac 930
 Db 482 AAGAGAGAGGTGAGGACTTCCCTATTCATATCTATGATGATTCATGAGAAATTMAAC 541
 QY 931 ttgaaagctactgtgagcgtgtgtatgtccctccagacataactgtgtccatgtgag 990
 Db 542 TTGAAAGGTCTACTCGAGCTGTATGATGATGATGATGATGATGATGATGATGAT 601
 QY 991 gagaagtcacccaagaacttcaagaagcgttcaagaagaactatttctgttctcccaat 1050
 Db 602 GAGAAGTCCCAAGAACTTCAAAAGAGGCGTCAAAAGCAATATTCTTCCCAAT 661
 QY 1051 gtcttcataagcagtaagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1107
 Db 662 GTCTTCATAGCCAGTAACTGCTGGGTGGGTATTAACCTTCCCTGAGGTCCAGGCTGCA 721
 QY 1108 --gtgacactcaactgtcatgaa--gaattgtccagaagctcagtgccgtggaata-c 1161
 Db 722 AGCTTGACCTTCAACTGCATGAGAAAGACTTGGCTCAGAGGCTGCGGTGAATFAC 781
 QY 1162 ttctcgaatacaatgtggg-----acgaacttccataaagaagcaatgag--agaatgt 1214
 Db 782 TTCCGAAATACCTTGTGGAAACCGGAGCTTTCATTAAGAGCAATGACAGAGATGGT 841
 QY 1215 ccaagctcagaagtgtgaatgag 1241
 Db 842 CCAGGCTCAAGATGTGACACGGGG 868

 RESULT 8
 B1765410 766 bp mRNA linear EST 25-SEP-2001
 LOCUS B1765410
 DEFINITION 603050445F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5190611 5',
 mRNA sequence.
 ACCESSION B1765410
 VERSION B1765410.1 GI:15756988

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 766)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLNL1476 row: 1 column: 12
High quality sequence stop: 766.
Location/Qualifiers
1. 766
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5190611"
/clone_lib="NIH_MGC_116"
/lab_host="DH10B"
/note="Organ: pooled colon, kidney, stomach; Vector: PCMV-SPORT6; Site: 1: NotI; Site: 2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH_MGC Library."

BASE COUNT 199 a 184 c 192 g 190 t 1 others

ORIGIN

Query Match 30.4%; Score 704.8; DR 10; Length 766;
Best Local Similarity 98.4%; Pred. No. 6.3e-194;
Matches 753; Conservative 0; Mismatches 8; Indels 4; Gaps 4;

QY 316 agaaagatcaaggagatggtgttccattgtgtccaggaacacccgcatcttact 375
DB 1 AGAAGATCAGGNGACAGG--TGTTACATTGCTGCCAGGACCCGACGCTTCAGG 59
QY 376 tggaaacagaatacagcctgtgaagatcatccctaaagcagagagaactactaaag 435
DB 60 GGGAAACAGAAATCAGCGCTGTGAAGAGATCCCTAAGACAGAGAGACTCTAAG 119
QY 436 gattgtctctctccaccttccctgtgtcgtgtccacctgttcccttctgtgagc 455
DB 120 GAGTGTCTCTCTCCACCTTCCTGTGCTGCTCCACCTGTCTCCATTCTGTAGC 179
QY 496 atggttcaatggagaagacctgcacgtcattactgtgtggctctgtggctctatg 555
DB 180 ATGTTCAATGGAAGACACTCTGCCAGCTCATTTACTGTGGGCTCTGGGCTGTATATG 239
QY 556 ctgctggccactgtgtgtcgtgaacttcttcaagttgaagtgtgacttgacactg 615
DB 240 CTGCTGGCCACTGTGGCTGAAACTTTCTTCAAGTTGAAGTGTGACTGTGACCACTTG 299
QY 616 gttctggagtcagggaatctcaagccagctactgtgagaaatctgtataatttcctg 675
DB 300 GGTCTGGAGTCAGGAGATCTCAAGCCACTACTGTGAATATCTTGAATTTCTCTG 359
QY 676 aaattcccaagaagaagttatcaactgttcaagggtgacccggaaggagcagaaggca 735
DB 360 AAACCTTCACGAAAGAGTCTATCACTGTCAAGGGGTACCCGAGGGGACCAAGAGGCA 419

QY 736 gtgttcaggctatcttgaataaccctggaggtcaagaagaagcagagccttcaagac 795
DB 420 GTGCTTCAGGCTATTTCTGAATACCTGAGAGTCAGAGAGAGAGAGCCTTTCACAGAC 479
QY 796 acccaactactcttccctaccagaagactgtgacacttcaaggtgaaggaattcata 855
DB 480 ACCCACTACTCTCTCCCTCACAGAGACTGTGACACTTCAAGGCTTAAGAGAGTTCA 539
QY 856 cagttccactgagcaagaagaggtggagttccctattgacactctatgtgtatcat 915
DB 540 CATTTCCTCATGAGCAAGAAGAGTGTGAGTTCCTTATTCATCTATGAGATTCAT 599
QY 916 gagaagattgaaaacttgaagagctactgagctgtgtatgtccctca-gaacaata 974
DB 600 GAGAAGATTGAAAACCTTGAAGAGCTACTGCGAGCTGTATGCCCTCAGAAACATATA 659
QY 975 ctgt-gtccatgtgagtgagaagatcccccgaacttcaag-agcggtcaagaagaatt 1032
DB 660 CTGTGTCCATGTGATGAGAGAGTCCCGACACTTTCAAAGAGCGGCTCAAGCAAT 719
QY 1033 attctgtctcccaaatgtcttcatagccaagtaagctgttcg 1077
DB 720 ATTTCTTGCTTCCCAATGATCTTCATAGCCAGTACCTGTGCGG 764

RESULT 9
BG386598 866 bp mRNA linear EST 12-MAR-2001
LOCUS 602454847F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583243 5',
DEFINITION mRNA sequence.
ACCESSION BG386598
VERSION BG386598.1 GI:13280044
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE 1 (bases 1 to 866)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC <http://mgc.nci.nih.gov/>.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLNL1306 row: n column: 12
High quality sequence stop: 744.
Location/Qualifiers
1. 866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583243"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 219 a 209 c 221 g 217 t

ORIGIN

Query Match 29.9%; Score 693.4; DB 10; Length 866;
 Best Local Similarity 97.9%; Pred. No. 1.4e-190;
 Matches 755; Conservative 0; Mismatches 11; Indels 5; Gaps 5;

442 gtcctccctccacccctccctgctggtcctccacactgtctcccatctgtgacgatggt 501
 |||||||
 Db 2 gttccctccctccacccctccctgctggtcctccacactgtctcccatctgtgacgatggt 501
 |||||||

502 catggaagaagactgctgcaagctgcatctactgtgtggtcgtggtcgtatgtcgtc 561
 |||||||
 Db 62 catggaagaagactgctgcaagctgcatctactgtgtggtcgtggtcgtatgtcgtc 121

562 gccactgtgtcctgtaaacattctcttcaagtgtaagtgtaactgtgaccactgtgctg 621
 |||||||
 Db 122 gccactgtgtcctgtaaacattctcttcaagtgtaagtgtaactgtgaccactgtgctg 181

622 ggtgccaggaagacttcaaaagcagtaactgttagaataatctgtataattcctgaaact 681
 |||||||
 Db 182 ggtgccaggaagacttcaaaagcagtaactgttagaataatctgtataattcctgaaact 240

682 ccaagcaagaagcttcatcaactgttcaaggtgtcaccggaggagccaaagaagcagtgctt 741
 |||||||
 Db 241 ccaagcaagaagcttcatcaactgttcaaggtgtcaccggaggagccaaagaagcagtgctt 300

742 caggtatattctgaataaactgtgaggttcaagaagaagcagagccttccacagacacccac 801
 |||||||
 Db 301 caggtatattctgaataaactgtgaggttcaagaagaagcagagccttccacagacacccac 360

802 taactctccctcaacagagactgtgagcaacttcaaggtcgaagaagaagtaactcaacttc 861
 |||||||
 Db 361 taactctccctcaacagagactgtgagcaacttcaaggtcgaagaagaagtaactcaacttc 420

862 ccactggaagaagaaggtgaggtgttccctattgcatctactatgtgttcatatagaag 921
 |||||||
 Db 421 ccactggaagaagaaggtgaggtgttccctattgcatctactatgtgttcatatagaag 480

922 attgaaactgtgaaagcgtactgcagctgtgtatgtccctccagacaatactatgtgtc 981
 |||||||
 Db 481 attgaaactgtgaaagcgtactgcagctgtgtatgtccctccagacaatactatgtgtc 540

982 catggtgataagaagttccccaagaacttccaagaagggcggtccaagcaatttctt-g 1040
 |||||||
 Db 541 catggtgataagaagttccccaagaacttccaagaagggcggtccaagcaatttctt-g 600

1041 ctccccaagtcttcatagcagcagtaagctgtgtcgtgtgttattgtcctctgtg-cca 1099
 |||||||
 Db 601 ctccccaagtcttcatagcagcagtaagctgtgtcgtgtgttattgtcctctgtg-cca 660

1100 ggggtgcaagctgcaactcaactgcatggaagactgtctccagagctca-9tgcgctgtgaa 1158
 |||||||
 Db 661 ggggtgcaagctgcaactcaactgcatggaagactgtctccagagctca-9tgcgctgtgaa 720

1159 tactctccgaatacatgt-gggaagagacttccctataaagaagcaatgacaga 1208
 |||||||
 Db 721 tactctccgaatacatgt-gggaagagacttccctataaagaagcaatgacaga 771

RESULT 10
 BG468641

LOCUS BG468641 731 bp mRNA linear EST 21-MAR-2001
 DEFINITION 603510243p1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:464467 5',
 mRNA sequence.
 ACCESSION BG468641
 VERSION BG468641.1 GI:13400911
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 731)
 AUTHORS NIH-MGC <http://mgi.mgi.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.lnl.gov>
 Plate: LNCM1416 row: e column: 12
 High quality sequence stop: 686.
 Location/Qualifiers
 1..731
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:464467"
 /clone_11b="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pOMB7; Site: 1: XhoI; Site: 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using 2AP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 214 a 157 c 183 g 177 t
 ORIGIN

Query Match 29.8%; Score 691.6; DB 10; Length 731;
 Best Local Similarity 98.4%; Pred. No. 4.3e-190;
 Matches 720; Conservative 0; Mismatches 9; Indels 3; Gaps 2;

568 gtggtcctgtaaacattctcttcaagtgtaagtgtaactgtgaccactgtggtcgtgaccc 627
 |||||||
 Db 2 gtggtcctgtaaacattctcttcaagtgtaagtgtaactgtgaccactgtggtcgtgaccc 61

628 aaggaatctcaagaagcagtaactgtggaataatctgtataattcctgaaactccaaga 687
 |||||||
 Db 62 aaggaatctcaagaagcagtaactgtggaataatctgtataattcctgaaactccaaga 121

688 aagaagctcatcaactgttcaaggggtcaccgagggagcgaagaagcagtgctcagact 747
 |||||||
 Db 122 aagaagctcatcaactgttcaaggggtcaccgagggagcgaagaagcagtgctcagact 181

748 attctgaataaactgtgaggttcaagaagaagcagagccttccacagacacccaactac 807
 |||||||
 Db 182 attctgaataaactgtgaggttcaagaagaagcagagccttccacagacacccaactac 241

808 tccctccacgaagactgtgagcaacttcaaggtcgaagaagaatcataagttccagact 867
 |||||||
 Db 242 tccctccacgaagactgtgagcaacttcaaggtcgaagaagaatcataagttccagact 301

868 agcaagaagaaggtgaggttccctattgtcatctactatgtgtgattcatagaagaattga 927
 |||||||
 Db 302 agcaagaagaaggtgaggttccctattgtcatctactatgtgtgattcatagaagaattga 361

928 aacttgaagaagctactgcagactgtgtatgtccctccagacaatactatgttccatgtg 987
 |||||||
 Db 362 aacttgaagaagctactgcagactgtgtatgtccctccagacaatactatgttccatgtg 421

988 gattgagaagttccccaagaacttcaagaagggcggtccaagaagcaatattctgtctcca 1047
 |||||||
 Db 422 gattgagaagttccccaagaacttcaagaagggcggtccaagaagcaatattctgtctcca 481

1048 aatgtcttcatagccaagtaagctgtgtgtgtgttatgctcctcctgaggtgcga 1107
 |||||||
 Db 482 aatgtcttcatagccaagtaagctgtgtgtgtgttatgctcctcctgaggtgcga 541

1108 gctgaactcaactgtaggaagactgtctccagagctcagtgccgtggaataactctctg 1167
 |||||||

Db 542 GCTGACCTCACTGATGAGAACTGCTCAGAGCTAGTCCCGNGAANAATCTTCTG 601
 Qy 1168 aatacatgttgagacgagcttccataaagaacatgcagagatggtccaggtctcaag 1227
 Db 602 AATACATGTGGAGAGCCTTTCATTAAGAGCAATGACAGATGCTCCAGG--CTCAG 659
 Qy 1228 atgttgatggaggaataagcatgagat-cagaggtccctccaaagacaagaacccg 1286
 Db 660 ATGTTGAAATGGAGGAATGACATGAGTACAGAGTAACATCATFAMGCAAAAAAGAACG 719
 Qy 1287 ctggaatatca 1298
 Db 720 TTGGAAATATTA 731

RESULT 11

LOCUS BG248304 702 bp mRNA linear EST 13-FEB-2001
 DEFINITION 602400340F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546017 5',
 mRNA sequence.

ACCESSION BG248304
 VERSION BG248304
 KEYWORDS BG248304.1 GI:12758119
 EST.

SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (Bases 1 to 702)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: NIH Intramural Sequencing Center.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
 Plate: LNCM1231 row: o column: 10
 High quality sequence stop: 702.
 Location/Qualifiers

FEATURES

source

1..702
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4546017"
 /clone_lib="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pOT87; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGAG(G). Site-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 186 a 163 c 175 g 178 t
 ORIGIN

Query Match 29.6%; Score 685.6; DB 10; Length 702;
 Best Local Similarity 99.4%; Pred. No. 2,4e-188;
 Matches 688; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 119 ttgtgtctggagagccctggtatctgtaatactactgttagtctgaaggaa 178
 Db 1 TTTTGTTCTGGGAAGCCCTGGGATTCTGCTAATACCTATGACTGTAGCTGCTGAAGGAA 60
 Qy 179 acagatgaagaacatgacccaagagctctcctgtaatgagaagaccagctacgct 238
 Db 61 ACAGATGAAGAACATGACCTCAAGAGCTTCTCTCATATGAGAAGACCAAGCTACGCCCT 120

Qy 239 ggcacaagataaagaagagcctgaaactgttccttgagacatctaatgaatcagaa 298
 Db 121 GGCAAAGATATTTAAAGGAGGCGCTGAAGACTGTTCTTGACATCTATGAAATGTCAAGAA 180
 Qy 299 ataccctttgaaaggttaagaagatcaaggagacatggtgttcaacatttgttcacagaa 358
 Db 181 ATACCTTTGGAGGGTAGAAGATCAGGGACATGGTTGTTCACTTTGCTGCCACAGAA 240
 Qy 359 caccgcagctctcaacttggaacaagatacagccctgttgaagatcaatccctaaagcag 418
 Db 241 CACCGCCAGTCTTACATTGGAAACAGATCACGCGCTGTGAAGAGATCATTCCCTAAGCAG 300
 Qy 419 gagaagaagactaaagagatgtgtctctccactccactccctgtgtcgtccacatg 478
 Db 301 GAGAGAAGCTACTAAAGAGATTGTCTCTCCACCTCCCTGCTGCGGTCTCCACTG 360
 Qy 479 tctccactctgtgagatgttcaatgagaagagacttcagcgctgcatctattgtggg 538
 Db 361 TCTCCCAATTGTGACATGTTCAATGAAGAGACTCTGCAGCTGCATTAAGTGTGGG 420
 Qy 539 ctctggcctgatatgtctgtgccaactgtgctctgaactcttctcaagttgaagt 598
 Db 421 CTCTGGGCTGCTATATGCTGCTGCGCACTGCGCTCGAAGCTTTCTTACAGTTGAGT 480
 Qy 599 gtgacttcgacacttggtgtcgtgagtcaggaatctcaagccagctactgaagata 658
 Db 481 GTGACTCTGACCACTTGCTGGAGTCCAGGGAATCTCAACACCACTGATGAGAAAT 540
 Qy 659 tctgtataattcttcgaactccagcaagaggttctcaactgttcagggttacc 718
 Db 541 TCTTGTAATATTCCTGAACCTTCACGAAAGAGGTATCAACTGTTCAAGGGGTACCC 600
 Qy 719 gaaggagacaaagagcagtgcttcaagcttatctgaataaccttgaagaagaagc 778
 Db 601 GAGGGACCAAGAGCGCTGCTTCAAGCTATGCTGGAATACCTGAGAGTCAAGAGAGC 660
 Qy 779 gagaagccttcaagaagaccactactctcc 810
 Db 661 GAGAGCCTTTCACAGAAACCATTAACGTCTCC 692

RESULT 12

LOCUS BG386247 920 bp mRNA linear EST 12-MAR-2001
 DEFINITION 602454539F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583984 5',
 mRNA sequence.

ACCESSION BG386247
 VERSION BG386247
 KEYWORDS BG386247.1 GI:13279693
 EST.

SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (Bases 1 to 920)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LNCM308 row: m column: 09
 High quality sequence stop: 749.
 Location/Qualifiers

FEATURES

source

1..920
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4583984"

/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT	249 a	210 c	238 g	223 t
ORIGIN				
Query Match	28.7%	Score 666.2;	DB 10;	Length 920;
Best Local Similarity	93.3%	Pred. No. 1.2e-182;		
Matches 751; Conservative	0;	Mismatches 48;	Indels 6;	Gaps 5;
Db	174	gggaacagatgaagaacatgaccccaaggagcttcctgcatgagaagacaagctga	233	
Qy	2	gggaacagatgaagaacatgaccccaaggagcttcctgcatgagaagacaagctga	61	
Db	234	cgctggaagaagataataagaagagcctgaactgttccttgacatctatgaatgc	293	
Qy	62	cgctggaagaagataataagaagagcctgaactgttccttgacatctatgaatgc	121	
Db	294	agaataaccttggaggaggttaagaagatcaggagacatgtttcacaatttgcctga	353	
Qy	122	agaaatattcc-tttggagaggtttagaaatcaggagacatgtttcacaatttgcctga	180	
Db	354	cggaaacacgcagctcttacttgaagaacagaatcacgccttgaagaagatcacctca	413	
Qy	181	cggaaacacgcagctcttacttgaagaacagaatcacgccttgaagaagatcacctca	240	
Db	414	agcagagagaagaagctactaaaggatgtgtccctccacacttccctgtgtcgtcc	473	
Qy	241	agcagagagaagaagctactaaaggatgtgtccctccacacttccctgtgtcgtcc	300	
Db	474	accttcccacacttctgagcagatgttcaatgaagaagagctctgcagctgactt	533	
Qy	301	accttcccacacttctgagcagatgttcaatgaagaagagctctgcagctgactt	360	
Db	534	gtgagctctgagctgactatgctgctgctgagcactgtgctgacacttcttcaagtt	593	
Qy	361	gtgagctctgagctgactatgctgctgctgagcactgtgctgacacttcttcaagtt	420	
Db	594	gaagtgtgactctgacacacttgagctgagatcgaaggaaatcgaagccagctag	653	
Qy	421	gaagtgtgactctgacacacttgagctgagatcgaaggaaatcgaagccagctag	480	
Db	654	gaatattctgtataatcttcctgaagaactctgcgaagaaggagctatcaactgttcagggt	713	
Qy	481	gaatattctgtataatcttcctgaagaactctgcgaagaaggagctatcaactgttcagggt	540	
Db	714	c-acccgagggagacaaagagcagtgctcagagctatctgaataaactgagagcaaga	772	
Qy	541	caacccgagggagacaaagagcagtgctcagagctatctgaataaactgagagcaaga	600	
Db	773	agaagcagagagccttcaacagacacacacacttccctcaacagagactgtgacact	832	
Qy	601	agaagcagagagccttcaacagacacacacacttccctcaacagagactgtgacact	660	
Db	833	t-caggagctgaagaagaatcatatagcttccacacgagcaagaagaagtgagttccct	891	
Qy	661	tcacagagctgaagaagaatcatatagcttccacacgagcaagaagaagtgagttccct	720	
Db	892	attgacatct-ctatggtgattcatcgagaaga--ctgaagaacttgaagaagctactgca	948	
Qy	721	attgacatctctctatggtgattcatcgagaaga--ctgaagaacttgaagaagctactgca	780	
Db	949	gctgtgatagtccctcagaacatat 973		

Db 781 ACTGGGTTGGCCCTGAGAACATATCT 805

RESULT 13
LOCUS BG386293
DEFINITION BG386293 871 bp mRNA linear EST 12-MAR-2001
602455486p1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583948 5',
RNA sequence.
ACCESSION BG386293
VERSION BG386293.1 GI:13279739
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 871)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LUCM1308 row: k column: 21
High quality sequence stop: 736.
Location/Qualifiers
1..871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583948"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT	231 a	212 c	215 g	213 t
ORIGIN				
Query Match	28.2%	Score 653;	DB 10;	Length 871;
Best Local Similarity	95.2%	Pred. No. 8.5e-179;		
Matches 793; Conservative	0;	Mismatches 25;	Indels 15;	Gaps 11;
Db	64	ctctaagtcacggaggaactgccccttgacttctgagcctgaccccttactcagcagttt	123	
Qy	2	ctctaagtcacggaggaactgccccttgacttctgagcctgaccccttactcagcagttt	60	
Db	124	ttcttggaagcccttggaattctgctaaactactacactgtagtgctgaaaggaaacga	183	
Qy	61	ttcttggaagcccttggaattctgctaaactactacactgtagtgctgaaaggaaacga	120	
Db	184	tgaagaacatgaacctcaaggagccttcctgctcaataagaagaagacagctgagccctggcaa	243	
Qy	121	tgaagaacatgaacctcaaggagccttcctgctcaataagaagaagacagctgagccctggcaa	180	
Db	244	agataataaagagagccttgaagaactgttccttgagcacttatagaatgtcagaataatcc	303	
Qy	181	agatatttaagagagccttgaagaactgttccttgagcacttatagaatgtcagaataatcc	240	
Db	304	tttggaggggttagaagatcaggaggaatggtgtgtcaccattgtcgcacgaagacacg	363	
Qy	241	tttggaggggttagaagatcaggaggaatggtgtgtcaccattgtcgcacgaagacacg	300	

QY 364 ccagcttcctacgttggaacaaacaaacacgacctgtgaagagatcatccctaaagcagagag 423
 |||||||
 Db 301 CCAAGTCTTCACTTGGAACAGATCAAGCTTGTGGAAGATCATCCCTTAAGCAGAGAG 360
 |||||||
 QY 424 aagctactaaagatgtgtctctctccacccctccctgtgtctgtctccacgtctcc 483
 |||||||
 Db 361 AACCTACTAAAGATGTGTCTCTCCACCTTCCCTGTCTGTCTGTCTGTCTGTCTCC 420
 |||||||
 QY 484 cactctgtgacgtgtgtatcat--ggaagagactctgacgctgacatctattgtgtgtc 541
 |||||||
 Db 421 CATTCTGTGAGATGTTCAATCGGGAAGAGACTGTGCCAGTGCATTTACTTGTGGGCTC 480
 |||||||
 QY 542 tgggtgtctatgtctgtctgtccactgtgtctgtgaacattctcttaagttgaagtgt 601
 |||||||
 Db 481 TGGGCTGCTATATGCTGCTGCTGCCACTGTGCTGTGAATCTTCTTTCAGGTGGAAGTGT 540
 |||||||
 QY 602 actctgacacactgt 661
 |||||||
 Db 541 ACTCTGACCACTTGT 599
 |||||||
 QY 662 tgtatata-ttctctgaacttcacagc-aagaaggtctataa-ctgtctgaggtgtaccc 718
 |||||||
 Db 600 TGTATATTTTCTTGAACCTTCAAGCAAAAGAGGTCTATCACTGTTCAGGGGTCAACC 659
 |||||||
 QY 719 ---gaggggacacagagagcagtgctt--caggtctatctgaataacactgtgaagtc 772
 |||||
 Db 660 CGAGGGGACCAAGAGGCGAGTGTCTTCCAGGGTATCTGATTAACCTGGAGGTCAACAG 719
 |||||
 QY 773 agaaagagagagcctttcac-agacacacactactctctctcaccagagagctgtgagc 830
 |||||||
 Db 720 AGAAGCGAGGCGCTTTCACAGACACCCACTACTCTTCTTCAACAGAGCTGGAGCAAA 779
 |||||||
 QY 831 ctccaagcgtgaaagagagcttccacagctgtgacagaaagaaagagtgag 883
 |||||||
 Db 780 CTTCAGAGCTGAAGGAGACGTACACAGCTCCCATAGCAGCAAAAGAAAGTGG 832
 |||||||

RESULT 14
 BE617113 736 bp mRNA linear EST 20-OCT-2000
 LOCUS 601441649F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3845677 5',
 DEFINITION mRNA sequence.
 ACCESSION BE617113
 VERSION BE617113.1 GI:9888051
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 736)
 NIH-MGC http://mgs.nci.nih.gov/.
 National Institutes of Health, Human Genome Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapds-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: L1AM9557 row: b column: 14
 High quality sequence start: 9
 High quality sequence stop: 670.
 Location/Qualifiers
 1. /736
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3845677"
 /clone_1lb="NIH_MGC_65"
 /tissue_type="adenocarcinoma"

FEATURES

source

1. /736
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3845677"
 /clone_1lb="NIH_MGC_65"
 /tissue_type="adenocarcinoma"

/lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pCW-SPOF6; Site: 1; NOTI:
 Site: 2; Salt: cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.8 kb. Library constructed by Life
 Technologies."
 BASE COUNT 191 a 170 c 183 g 191 t 1 others
 ORIGIN

Query Match 28.1%; Score 651.8; DB 10; Length 736;
 Best Local Similarity 99.3%; Pred. No. 1,7e-178;
 Matches 675; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 86 ttgctacttgagccctcccttaccacagagctttgtcttggaagccctggattct 145
 |||||||
 Db 1 TTGCTACTTGTGAGCTGCGCCCTTACACAGAGTTTGTTGTTCTGTGGAAAGCCCTGGATTC 60
 |||||||
 QY 146 gctataactatacactgtgaagtgctgaagggaacagatgaagaacatgacctcaaggag 205
 |||||||
 Db 61 GCTAATACCTATACCTAGTACTGAGTGTGAGGAAACAGATGAACATGACCTCAAGGAG 120
 |||||||
 QY 206 ctctcgtctaatgaagagacacagcctgagcctggaagaagattaaagagagcctgaa 265
 |||||||
 Db 121 CTTCCTGTCAATGAGAGACCAAGCTGACGCTGTGGCAAAAGATATTAAGAGAGCTGAA 180
 |||||||
 QY 266 actgtctcttgagacatcttatagtatgcaagaataacctttgagaggttagaagatcag 325
 |||||||
 Db 181 ACTGTTCCCTTGAGACATCTTATGTAATGTCAAGAAATACC-TTGTGAGGGTTAAGAAATCAG 229
 |||||||
 QY 326 gggacatggtgtgtcacattgtctgccacggaaacccgcagctcttcaactgtgaacaga 385
 |||||||
 Db 240 GGGACATGGTGTTCACATTTGCTGCGCACGAAACACCGCAGCTTCACTTGGAAACAGA 299
 |||||||
 QY 386 atcagcctgtgtgaagaagatcatccctaaagcagagagaagctactcaagaagatgtgttc 445
 |||||||
 Db 300 ATACGCGCTTGTGAAGAGATCATCTTACACAGAGAGAAAGCTACTTAAGAGATTGTCTC 359
 |||||||
 QY 446 tccctcaacttccctgtgtctgtgtctcaccgtgtctccactctgtgtgaagatgttcaat 505
 |||||||
 Db 360 TCTCCACACTTCCCTGCTGCTGCTGCCACCTGTCTCCATCTGTGACGATGTTCAAT 419
 |||||||
 QY 506 ggaagagacctctgcagctgcatctactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 565
 |||||||
 Db 420 GGAAGAGACTCTCCACCTGCATTAATCTGTGGCTCTGGGCTCTGTATATCTCTGTG-CA 478
 |||||||
 QY 566 ctgtgtctgtgaacttcttctcaggtgtgaagtgtaactgtgacacactgtgtgtgtgt 625
 |||||||
 Db 479 CTGTGCTGTGAACACTTCTTTCAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 538
 |||||||
 QY 626 ccaagggaatcacaagcagactgtgaagaatatctgtataatctcctgaacactccag 685
 |||||||
 Db 539 CCAGGGAATCTCAAGGCACTACTGTAAGTAATCTGTATATTTCTGTAACCTCCAG 598
 |||||||
 QY 686 caaagaggtctatcaactgtgtcaggggtcacccgaagagggcgaagagggcaggtctcagg 745
 |||||||
 Db 599 CAAAGAGGTATATCAATGTTCAAGGGGTCAACCCGAGGAGCAAGAGGAGGTCTTCAGG 658
 |||||||
 QY 746 ctattctgaataacactgtgag 765
 |||||||
 Db 659 CTATTCTGATAATCACTGAG 678
 |||||||

RESULT 15

BG328280

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

602427256F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546813 5',
 mRNA sequence.
 BG328280 BG328280.1 GI:13134627
 EST.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 955)
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM1233 row: P column: 14
High quality sequence stop: 718.

FEATURES
Location/Qualifiers
1..955
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4546813"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAC(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 245 a 230 c 246 g 233 t 1 others
ORIGIN

Query Match 28.1%; Score 650.8; DB 10; Length 955;
Best Local Similarity 96.9%; Pred. No. 3.9e-178;
Matches 717; Conservative 0; Mismatches 17; Indels 6; Gaps 5;

QY 64 ccttaagtcagggaactgccccttgctactgtgacgtgccccttaactcagcagttttg 123
DB 2 cttctagtcacgsggaactcccc-tgctacttgtagctgccccttactcagcag-
QY 124 tctctgggaagccctggagcttctgctaatactacatcacgttagtgctggaaggaacaga 183
DB 60 tttctgggaagccctggagcttctgctaatactacatcacgttagtgctggaaggaacaga 119
QY 184 tgaagaacatgacactcaagaagcttctcgtcaatgagaagcaaacagcctgtgcaa 243
DB 120 tgaagacatgacactcaagaagcttctcgtcaatgagaagcaaacagcctgtgcaa 179
QY 244 agatatlaagaagagcctgaaactgttctctggacatctatgaatgtcagaataacc 303
DB 180 agatatlaagaagagcctgaaactgttctctggacatctatgaatgtcagaataacc 239
QY 304 ttttgaagggttaagaagatcaggggacatggtgttcaacattgtgtgcagcaggaacag 363
DB 240 ttttgaagggttaagaagatcaggggacatggtgttcaacattgtgtgcagcaggaacag 299
QY 364 ccagcttcaacttggaacagaaacagacacgctgtggaagatcaccctcaagcagagag 423
DB 300 ccagcttcaacttggaacagaaacagacacgctgtggaagatcaccctcaagcagagag 359
QY 424 aagctactaagagatgtgtcctcctcgaactcctgtgtcgtgctcgaactgtctcc 483
DB 360 aagctactaagagatgtgtcctcctcgaactcctgtgtcgtgctcgaactgtctcc 419
QY 484 catctcgtacagatggttcaatggaagagacatgcgaactgcatctactgtgtgtgtg 543
DB 420 catctcgtacagatggttcaatggaagagacatgcgaactgcatctactgtgtgtgtg 479
QY 544 ggcctcatalatgctgtgcacactgtgtgtcgaacttctcagttgaagtgtgac 603

DB 480 ggcctcatalatgctgtgcacactgtgtgtcgaacttctcagttgaagtgtgac 539
QY 604 tctgaccacttggtgtcgtggaagatcctcaagccagacagctgtagaatacttg 663
DB 540 tctgaccacttggtgtcgtggaagatcctcaagccagacagctgtagaatacttg 599
QY 664 tataattcctg-aaacttccagcaagaggtctacactgttcaggggtcaccgag 722
DB 600 tataatttcttgaaacttccagcaagaggtctacactgttcaggggtcaccgag 657
QY 723 ggaaccaagagcagtgcttcaagctatctgaataacctggaaggtcaagaagcagaga 782
DB 658 ggaaccaagagcagtgcttcaagctatctgaataacctggaaggtc-agaagaagcagaga 716
QY 783 ggccttcacagacaccact 802
DB 717 gctttacagaAACCAATACT 736

Search completed: September 26, 2002, 07:00:01
Job time: 14156 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 01:53:55 : Search time 4629.8 Seconds
(without alignments)
1107.392 Million cell updates/sec

Title: US-09-874-390-1_COPY_1_245

Perfect score: 245
Sequence: 1 attaactgggtttccatt.....caagctgacgcctgccaag 245

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: GenEmbl:*
2: gb_ba:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
------------	-------------	-------	--------	----	-------------

1	245	100.0	174323	2	AC092754	AC092754 Homo sapi
2	245	100.0	187275	2	AC092755	AC092755 Homo sapi
3	176	71.8	2216	6	AX045249	AX045249 Sequence
4	176	71.8	2217	6	AX045251	AX045251 Sequence
5	176	71.8	2217	6	AX045253	AX045253 Sequence
6	58	23.7	2106	9	BC017032	BC017032 Homo sapi
7	37	15.1	101204	9	AC109761	AC109761 Rattus no
8	37	15.1	109101	9	AL136446	AL136446 Homo sapi
9	37	15.1	170695	2	AC068545	AC068545 Homo sapi
10	37	15.1	253217	2	AC016590	AC016590 Homo sapi
11	36.8	15.0	156733	9	AC069411	AC069411 Homo sapi
12	36.8	15.0	193295	2	AC073093	AC073093 Homo sapi
13	35.4	14.4	150650	2	AC021770	AC021770 Homo sapi
14	35.4	14.4	161733	2	AC013635	AC013635 Homo sapi
15	35.4	14.4	192696	9	AC009869	AC009869 Homo sapi
16	35	14.3	39882	9	AL356289	AL356289 Human DNA
17	35	14.3	59905	2	AC100985	AC100985 Mus muscu
18	35	14.3	98985	2	AL139427	AL139427 Homo sapi
19	35	14.3	125290	9	AC004986	AC004986 Homo sapi
20	35	14.3	146255	2	AC084170	AC084170 Homo sapi
21	35	14.3	178984	2	AC018540	AC018540 Homo sapi
22	35	14.3	212712	2	AC063940	AC063940 Homo sapi
23	34.8	14.2	199211	2	AC097054	AC097054 Rattus no
24	34.6	14.1	193564	2	AC084779	AC084779 Mus muscu
25	34.6	14.1	235693	2	AC084780	AC084780 Mus muscu
26	34.4	14.0	104115	2	AC022426	AC022426 Homo sapi
27	34.4	14.0	105565	2	AC104103	AC104103 Mus muscu
28	34.4	14.0	117735	9	AC022440	AC022440 Homo sapi
29	34.4	14.0	148103	9	AL445232	AL445232 Human DNA
30	34.4	14.0	201129	2	AC011358	AC011358 Homo sapi
31	34.2	14.0	100834	9	AL160411	AL160411 Human DNA
32	34.2	14.0	251492	2	AL663077	AL663077 Mus muscu
33	34.2	14.0	292494	2	HS313817A	HS313817A Homo sapi
34	33.8	13.8	65243	2	AC104669	AC104669 Homo sapi
35	33.8	13.8	162932	2	AC018516	AC018516 Homo sapi
36	33.6	13.7	97225	2	AC099417	AC099417 Rattus no
37	33.6	13.7	105891	9	AC004850	AC004850 Homo sapi
38	33.6	13.7	121684	2	AL354655	AL354655 Homo sapi
39	33.6	13.7	125787	9	AL645820	AL645820 Human DNA
40	33.6	13.7	125787	9	HS113817A	HS113817A Homo sapi
41	33.6	13.7	172000	1	AF000005	AF000005 Pyrococcu
42	33.6	13.7	197620	9	AF001905	AF001905 Homo sapi
43	33.4	13.6	35822	9	AC004609	AC004609 Homo sapi
44	33.4	13.6	105040	9	AC011229	AC011229 Homo sapi
45	33.4	13.6	148005	2	AC024944	AC024944 Homo sapi

ALIGNMENTS

RESULT 1
AC092754 LOCUS AC092754 Homo sapiens chromosome 15 clone RP11-112N19 map 15q22, ***
DEFINITION AC092754 SEQUENCING IN PROGRESS ***, 3 ordered pieces.
ACCESSION AC092754
VERSION AC092754.3 GI:18249988
KEYWORDS HTG, HTGS_PHASE2, HTGS_ACTIVEPIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 174323)
Rowen,J., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Burke,J., Dots,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R., Pate,D., and Hood,L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 174323)
Rowen,J., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Burke,J., Dots,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R., Pate,D., and Hood,L.
AUTHORS Direct Submission
TITLE

JOURNAL

Submitted (26-JUL-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
On Jan 19, 2002 this sequence version replaced gi:15022676.

QY 241 caaag 245
|||||
Db 173121 CAAAG 173125

COMMENT

Center: Multimegabase Sequencing Center
Center code: UMNMC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leewen@systemsbiology.org
----- Summary Statistics -----
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-terminator Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Insert size: -----; agarose-1p
Quality coverage: ---x in Q20 bases; sum-of-contigs

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and the accession number will be preserved.

* 1 99776: contig of 99776 bp in length
* 99777 99876: gap of unknown length
* 99877 172125: contig of 72249 bp in length
* 172126 172225: gap of unknown length
* 172226 174323: contig of 2098 bp in length.

FEATURES

SOURCE

1.174323
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q22"
/clone_lib="RP1-112N19"
/note="This clone overlaps RP11-99L18 and RP11-361D15"
BASE COUNT 46623 a 36456 c 37881 g 53163 t 200 others
ORIGIN

Query Match 100.0%; Score 245; DB 2; Length 174323;
Best Local Similarity 100.0%; Pred. No. 3.7e-70;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attactgaggttcctattatcatcctcgcattactcctcgtgtagagccct 60
|||||
Db 172881 ATTAACTGCGTTTCCATTATATCTATCCTCGCATTTACTTCTGAGTAGAGCCCT 172940
QY 61 tctcttaagtaacggaactcctctgctactgtgacgcctccttactcaagcatt 120
|||||
Db 172941 TCTCTTAAGTACGAGGAACTGCCCTTGCTACTTGTCAGCTGCCCTTTACACAGATT 173000
QY 121 ttgtctggggaagccctgggattctgtctaataactataactacgtgagtgctaaaggaac 180
|||||
Db 173001 TTGTTCTGGGAAGCCCTGGGATTCTGCTAAATCACTATCACTGAGTGTCTGAAGGAAC 173060
QY 181 agatgaagaataacactcaagagagctcctcgtctaataagagaagcaagcgcagccgg 240
|||||
Db 173061 AGATGAAGAATCACTCACTCAAGAGCTTCTGTCATATGAGAAGCAAGCTGACGCTGG 173120

RESULT 2
AC092755 187275 bp DNA linear HTG 19-JAN-2002
LOCUS Homo sapiens chromosome 15 clone RP11-361D15 map 15q22, ***
DEFINITION SEQUENCING IN PROGRESS ***; 2 ordered pieces.
AC092755 AC022480
AC092755.3 GI:18249989
VERSION HTG: HTGS_PHASE2; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Submitted (26-JUL-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
On Jan 19, 2002 this sequence version replaced gi:15022677.

Center: Multimegabase Sequencing Center
Center code: UMNMC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leewen@systemsbiology.org
----- Summary Statistics -----
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-terminator Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Insert size: -----; agarose-1p
Quality coverage: ---x in Q20 bases; sum-of-contigs

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and the accession number will be preserved.

* 1 32680: contig of 32680 bp in length
* 32681 32781: gap of unknown length
* 32781 187275: contig of 154495 bp in length.

FEATURES

SOURCE

1.187275
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"

/map="15q22"
/clone="RP11-361D15"
/clone.lib="RP11 human BAC library 11"
/note="This clone overlaps RP11-112N19 and RP11-568G20"
BASE COUNT 53688 a 38780 c 38083 g 56624 t 100 others
ORIGIN

Query Match 100.0%; Score 245; DB 2; Length 187275;
Best Local Similarity 100.0%; Pred. No. 3.8e-70;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attactggttttctatctatctatctctctgcattactctctgagtcagagcctct 60
Db 33616 ATTAACTGGGTTTCCATTATCATCTCTCGATTACTCTCTGTAGTCAGACCTCT 33675

QY 61 tctcttaagtcacgggaagcctgcttctacttctgagcctgccccttaactcaagttt 120
Db 33676 TCTCTTAAGTCACGGGAACGCTTCTCTACTGTGACGCTTACTACAGCAGTTT 33735

QY 121 ttgtctgggaagccctggagttctgtcaataactatctactgtgagtcgaaggaaac 180
Db 33736 TTGTTCTGGGAAGCCCTGGGATTCTGCTAATACCTATCATCTAGTCTGAAGGGAAC 33795

QY 181 agatgaagaacatgacctcaagagctctccgtcaatgagaagacagcctgagcctcg 240
Db 33796 AGATGAAGAACATGACCTCAAGAGCTCTCTGCTAATGAGAAGCAAGCTGACCCCTGG 33855

QY 241 caaag 245
Db 33856 CAAAG 33860

RESULT 3
AX045249
LOCUS AX045249 2216 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 1 from Patent WO0066727.
ACCESSION AX045249
VERSION AX045249.1 GI:11343799
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2216)
AUTHORS Adol.f.G., Heider.K.H. and Sommergruber.W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 0066727-A 1 09-NOV-2000;
Boehringer Ingelheim International GmbH (DE)

FEATURES
source location/Qualifiers
1..2216
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..426
427..1743
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17334.1"
/db_xref="GI:11343800"

5'UTR
CDS
3'UTR
BASE COUNT 600 a 499 c 547 g 570 t
ORIGIN

Query Match 71.8%; Score 176; DB 6; Length 2216;

Best Local Similarity 100.0%; Pred. No. 2.3e-47;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 gtcaaggaactgcccctgtactgtgactgccccttaactcaagcagttttgtctgg 129
Db 1 GTCACGGGAAGTGGCCCTGTACTGTGACCTGCGCCCTTACTACACAGTTTGTCTGG 60

QY 130 gaagccctggattctgttaatactacactgtatgtctggaaggaacagatgaaga 189
Db 61 GAAGCCCTGGGATTGTGCTAATACCTACTGTAGTGCTGAAGGAACAGATGAAGA 120

QY 190 aacatgacctcaagagagctctgtcaatgagaagaccaaagtcagcctggcaaag 245
Db 121 ACATGACCTCAAGAGAGCTTCCGTCAATGAGAAGCAAGCTGACGCTGGCAAG 176

RESULT 4
AX045251
LOCUS AX045251 2217 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 3 from Patent WO0066727.
ACCESSION AX045251
VERSION AX045251.1 GI:11343801
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2217)
AUTHORS Adol.f.G., Heider.K.H. and Sommergruber.W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 0066727-A 3 09-NOV-2000;
Boehringer Ingelheim International GmbH (DE)

FEATURES
source location/Qualifiers
1..2217
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..844
845..1744
1745..2217
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17336.1"
/db_xref="GI:11343803"

CDS
3'UTR
BASE COUNT 600 a 498 c 547 g 572 t
ORIGIN

Query Match 71.8%; Score 176; DB 6; Length 2217;
Best Local Similarity 100.0%; Pred. No. 2.3e-47;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 gtcaaggaactgcccctgtactgtgactgccccttaactcaagcagttttgtctgg 129
Db 1 GTCACGGGAAGTGGCCCTGTACTGTGACCTGCGCCCTTACTACACAGTTTGTCTGG 60
QY 130 gaagccctggattctgttaatactacactgtatgtctggaaggaacagatgaaga 189

```

Db      61 GAAGCCCTGGGATTCTGCTAATACCTATCCTAGTCTGGAAGGGAACAGATGAAGA 120
Qy      190 acatgaacctaaagagcttcctgcaatgagaagacaaactgaagcgtgcaaa 245
        |||||||
Db      121 ACATGACCTCAAGAGCTTCTCTCATATGAGAACCAAGCTGACGCTGCAAAAG 176

RESULT  5
LOCUS   AX045253
DEFINITION Sequence 5 from Patent WO0066727.
ACCESSION AX045253
VERSION  AX045253.1 GI:11343804
KEYWORDS
SOURCE  human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
  1 (bases 1 to 2217)
AUTHORS  Adolf,G., Heider,K.H. and Sommergruber,W.
TITLE    Tumour-associated antigen
JOURNAL  Patent: WO 0066727-A 5 09-NOV-2000.
FEATURES
  source
    Location/Qualifiers
      1..2217
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        1..844
        845..1744
        /note="unnamed protein product"
        /codon_start=1
        /protein_id="CAC17337.1"
        /db_xref="GI:11343805"
        /translation="MRRLKGLGCELCMPRLRTYVSMMKKSPETFEKAVKALISCF
          PVVFASKIVRVVYASNSRVOADLNCMEDLQSSVPRKYELNMGCPDPRKSAEMVO
          ALKMLNGRSMSEVPKPKHETRMKTHFEVYRDITLITNKKDPPRYNLMTFTGNAYI
          VASRDVQVHLNPKNSQQLIEWKDYSDDEHLMATLQARWMPGSPVPHPRYDISDM
          TSIRARLVKQWQHGIDDKGAPYAPCGIHQRAICVYGADLNMWLNHHLANKFDPK
          VDDNALQCLEEYLRKAYIGTEL"
          1745..2217
BASE COUNT 600 a 498 c 547 g 572 t
ORIGIN
3'UTR
BASE COUNT 600 a 498 c 547 g 572 t
ORIGIN

Query Match      71.8%; Score 176; DB 6; Length 2217;
Best Local Similarity 100.0%; Pred. No. 2.3e-47;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      70 gtaacgggaactgccttgctactgtgacctgcaccttaactgaagctttgtctcg 129
        |||||||
Db      1 GTCACGGGAACCTCCCTGCTACTGCTGACCTGCCCTTACTCAGCAGTTTGTCTCG 60

Qy      130 gaagccctggagattctgataaccataccctgtagtgcctgaagggaacagatgaaga 189
        |||||||
Db      61 GAAGCCCTGGGATTCTGCTAATACCTATCCTAGTCTGGAAGGGAACAGATGAAGA 120

Qy      190 acatgaacctaaagagcttcctgcaatgagaagacaaactgaagcgtgcaaa 245
        |||||||
Db      121 ACATGACCTCAAGAGCTTCTCTCATATGAGAACCAAGCTGACGCTGCAAAAG 176

RESULT  6
LOCUS   BC017032
DEFINITION Homo sapiens, glucosaminyl (N-acetyl) transferase 3, mucin type,
ACCESSION BC017032
VERSION  BC017032.1 GI:16877561
KEYWORDS  MGC.
SOURCE  human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE 1 (bases 1 to 2106)
AUTHORS  Strausberg,R.
TITLE    Direct Submission
JOURNAL  Submitted (05-NOV-2001) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK  NIH-MGC Project URL: http://mgc.ncl.nih.gov
CONTACT  MGC help desk
EMAIL    cgaabs-remail.nih.gov
Tissue Procurement: AFCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
          (Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
          R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: TRAK Plate: 20 Row: n Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4758421.
Location/Qualifiers
  1..2106
    /organism="Homo sapiens"
    /db_xref="LocusID:9245"
    /db_xref="taxon:9606"
    /clone="MGC:9086 IMAGE:3851937"
    /tissue_type="Colon, adenocarcinoma"
    /clone_lib="NIH_MGC_65"
    /lab_host="DH10B"
    /note="vector: pCMV-SPORT6"
    309..1625
    /product="glucosaminyl (N-acetyl) transferase 3, mucin
    type"
    /protein_id="AAH17032.1"
    /db_xref="GI:16877562"
    /translation="MYQWRKLCQHLVLAAGCYMLATYALKSPRKDKSDSHLGIES
      RESQSOYCRNLILNFKLPKRSINSGVTRGDQAVLADAILNNLEVKKKEPFDTH
      YLSLRDCEHFAERKFIQPLSKSEVEPRVIAVSMYHRIENFERLRAVAPONTY
      CYHVDKSPETFEKAVKAIISCFPNVFIASKLVRVYASMSRVOADLNCMEDLQSSV
      PKWYFNTGTPDKPIKSNAMVOALKMLNGRSMSEVPKPKHETRMKTHFEVYRDITL
      HTFNKKKDDPPRYNLMTFTGNAYIVASRDVQVHLNPKNSQQLIEWKDYSDDEHLM
      TQARWMPGSPVPHPRYDISDMTSIRARLVKQWQHGIDDKGAPYAPCGIHQRAICV
      YGAGDLNMWLNHHLANKFDPKVDNALQCLEEYLRKAYIGTEL"
BASE COUNT 582 a 471 c 516 g 537 t
ORIGIN

Query Match      23.7%; Score 58; DB 9; Length 2106;
Best Local Similarity 100.0%; Pred. No. 3.7e-08;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      188 gaacatgacctaaagagcttcctgcaatgagaagacaaactgaagcgtgcaaa 245
        |||||||
Db      1 GAACATGACCTCAAGAGCTTCTCTCATATGAGAACCAAGCTGACGCTGCAAAAG 58

RESULT  7
LOCUS   AC109761
DEFINITION Rattus norvegicus clone CH230-310K9, *** SEQUENCING IN PROGRESS
ACCESSION AC109761
VERSION  AC109761.1 GI:18580874
KEYWORDS  HTG: HTGS_PHASE1.
SOURCE  Norway rat.
ORGANISM Rattus norvegicus

```


REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 101204)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Albrooks, S.L., Amaral, H.C., Are, J.R., Ayala, M., Banks, T.,
Barbata, J., Benton, J., Blum, K., Blum, K., Blum, K., Blum, K.,
Bouck, J., Bowe, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
Burch, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Daya, M., Davis, C., Davy-Carroll, L., Dederich, D.A.,
DeLaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, U., Foster, P., Frantz, P.,
Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhley, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Mosier, M., Neal, D., Newton, J., Newton, N., Nguyen, N., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okunolu, G.,
Orduy, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojokokan, I., Rolfe, M., Ruiz, S., Saverly, G.,
Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Syatke, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Taneay, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Uman, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wall, R.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watson, R., Wall, R.,
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

Unpublished

2 (bases 1 to 101204)

Worley, K.C.

Direct Submission

Submitted (07-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GQVI
Center clone name: CH230-310K9
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 82934 bases at least Q40
Consensus quality: 87376 bases at least Q30
Consensus quality: 90911 bases at least Q20
Estimated insert size: 83906; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 42 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 5456: contig of 5456 bp in length
* 5457: gap of unknown length
* 5557: contig of 5151 bp in length
* 10707: gap of unknown length
* 10808: contig of 5472 bp in length
* 16279: gap of unknown length
* 16379: gap of unknown length
* 16380: contig of 4749 bp in length
* 21128: gap of unknown length
* 21228: contig of 3181 bp in length
* 24409: gap of unknown length
* 24509: contig of 3235 bp in length
* 27744: gap of unknown length
* 27745: contig of 2866 bp in length
* 30710: gap of unknown length
* 30711: contig of 4047 bp in length
* 30811: gap of unknown length
* 34857: gap of unknown length
* 34957: contig of 2624 bp in length
* 37581: gap of unknown length
* 37582: contig of 2907 bp in length
* 40588: gap of unknown length
* 40589: contig of 2512 bp in length
* 40689: gap of unknown length
* 43200: contig of 4556 bp in length
* 43301: gap of unknown length
* 47856: contig of 2337 bp in length
* 47956: gap of unknown length
* 50293: contig of 2359 bp in length
* 50394: gap of unknown length
* 52993: gap of unknown length
* 52993: contig of 2359 bp in length
* 53093: gap of unknown length
* 53093: contig of 3463 bp in length
* 56555: gap of unknown length
* 56556: contig of 2588 bp in length
* 59193: gap of unknown length
* 59293: contig of 2110 bp in length
* 59294: gap of unknown length
* 61403: contig of 2110 bp in length
* 61404: gap of unknown length
* 61503: contig of 2289 bp in length
* 63793: gap of unknown length
* 63892: contig of 1126 bp in length
* 63893: gap of unknown length
* 65018: contig of 2362 bp in length
* 65119: gap of unknown length
* 67501: contig of 1360 bp in length
* 67601: gap of unknown length
* 68960: contig of 1360 bp in length
* 68961: gap of unknown length
* 69060: contig of 2354 bp in length
* 71414: gap of unknown length
* 71415: contig of 1702 bp in length
* 71515: gap of unknown length
* 73217: contig of 1514 bp in length
* 74830: gap of unknown length
* 74831: contig of 1340 bp in length
* 74931: gap of unknown length
* 76271: contig of 1424 bp in length
* 76371: gap of unknown length
* 77795: contig of 1156 bp in length
* 77895: gap of unknown length
* 79051: contig of 1460 bp in length
* 79151: gap of unknown length
* 80611: contig of 1165 bp in length
* 80711: gap of unknown length
* 81875: contig of 1570 bp in length
* 81976: gap of unknown length
* 83545: contig of 1399 bp in length
* 83646: gap of unknown length
* 85044: contig of 1658 bp in length
* 85145: gap of unknown length
* 86842: contig of 1658 bp in length
* 86843: gap of unknown length
* 88741: contig of 1799 bp in length

*	88742	88841: gap of unknown length
*	88842	89872: contig of 1031 bp in length
*	89873	89972: gap of unknown length
*	89973	91217: contig of 1245 bp in length
*	91218	91317: gap of unknown length
*	91318	92973: contig of 1656 bp in length
*	92974	93073: gap of unknown length
*	93074	94267: contig of 1194 bp in length
*	94268	94367: gap of unknown length
*	94368	95652: contig of 1285 bp in length
*	95653	95752: gap of unknown length
*	95753	96962: contig of 1210 bp in length
*	96963	97062: gap of unknown length
*	97063	98694: contig of 1632 bp in length
*	98695	98794: gap of unknown length
*	98795	100063: contig of 1269 bp in length
*	100064	100163: gap of unknown length
*	100164	101204: contig of 1041 bp in length

FEATURES

source

```
/organism="Rattus norvegicus"  
/db_xref="taxon:10116"
```

[illegible]

Query Match	15.1%	Score 37	DB 2	Length 101204
Best Local Similarity	52.28%	Pred. No. 0.47%		
Matches 82	Conservative	0	Mismatches 75	Indels 0
				Gaps 0

QY 9 ggttttcctattatctatccctctcgcatcttctctgagtcagagcccttctctta 68

Db 1.980 GCTCTTGCTTCCCTCCTCTCTGCTGCTCTTGTCCCTTCTCTGTCTCTCTCTA 1921

69 agtcacgggaactgcccctgctaattgacccctgcccttactcagcagttttgtctcg 128

DD 1720 CCCCCCAGCIIICCCACAGAGCCCGGCCCTTGGTTCCCTCCCCTC 1801

Db 1860 TTCTACTCTTCTTCTCTCATTTAAACCTTTCCACGTTG 1824

RESULT 8
AL136446/c

DEFINITION Human DNA sequence from clone RP3-446119 on chromosome 6, complete

ACCESSION	AL136446
VERSION	AL136446.14
	GT:11967455

NETWORKS
SOURCE

0
1
2
3
4
5
6
7
8
9

REFERENCE

JOURNAL

COMMENT

On Dec 22, 2000 this sequence version replaced g1:118/5594. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is known to contain repeats, the repeat elements are annotated as such.

ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/Help/Chrb> R3J-446119 is from the library RPCR-3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/dacpac/home.htm>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone R3J-446119 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone R3J-446119 is at 1 in this sequence. The true left end of clone RP1-12409 is at 109002 in this sequence. The true right end of clone RP11-34662 is at 35550 in this sequence.

Location/Qualifiers

1..109101

```
/db_xref="taxon:9606"  
/chromosome="6"
```

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeal_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

[illegible]

repeat_region

1

3
0
5
0
3
+

3
0
2
..
0
3

repeat_region

repeated-regression

```
repeat_region /note="THEIB-INTERNAL repeat: matches 1. .433 of consensus"
20697. .21060
/note="THEIB repeat: matches 1. .364 of consensus"
repeat_region 21164. .21371
/note="MIR repeat: matches 34. .251 of consensus"
21581. .21916
/note="AluYo repeat: matches 1. .312 of consensus"
24571. .24662
/note="L2 repeat: matches 2334. .2428 of consensus"
repeat_region 26124. .26422
/note="AluY repeat: matches 1. .299 of consensus"
28340. .28740
/note="L2 repeat: matches 1535. .1713 of consensus"
repeat_region 28741. .29040
/note="AluYo repeat: matches 1. .297 of consensus"
29041. .29493
/note="L2 repeat: matches 990. .1535 of consensus"
repeat_region 29493. .29999
/note="L2 repeat: matches 1309. .1860 of consensus"
repeat_region 32968. .33150
/note="MIR repeat: matches 65. .262 of consensus"
35906. .36171
/note="LIME repeat: matches 937. .1211 of consensus"
repeat_region 36608. .38632
/note="L2 repeat: matches 202. .2471 of consensus"
repeat_region 38708. .38899
/note="2 copies 96 mer 80% conserved"
39435. .39762
/note="MER58B repeat: matches 1. .340 of consensus"
repeat_region 40470. .40495
/note="13 copies 2 mer aa 92% conserved"
41247. .41326
/note="MAD1 repeat: matches 1. .80 of consensus"
43677. .43997
/note="L2 repeat: matches 2402. .2708 of consensus"
48911. .49064
/note="L1MB1 repeat: matches 5628. .5776 of consensus"
49065. .49311
/note="L1PA13 repeat: matches 5909. .6155 of consensus"
49313. .49684
/note="L1PA10 repeat: matches 5774. .6158 of consensus"
50119. .50342
/note="L1MB1 repeat: matches 5935. .6170 of consensus"
51081. .51267
/note="MIR repeat: matches 46. .245 of consensus"
51273. .51316
/note="L1R25 repeat: matches 145. .190 of consensus"
52873. .52915
/note="MIR repeat: matches 211. .253 of consensus"
52874. .52926
/note="L2 repeat: matches 2694. .2746 of consensus"
56033. .56405
/note="MSTD repeat: matches 1. .394 of consensus"
56465. .56555
/note="L2 repeat: matches 2579. .2671 of consensus"
56556. .56875
/note="MER2 repeat: matches 1. .345 of consensus"
56876. .57207
/note="L2 repeat: matches 2262. .2579 of consensus"
58300. .58567
/note="AluY repeat: matches 24. .303 of consensus"
60945. .61032
/note="MIR repeat: matches 61. .145 of consensus"
61490. .61596
/note="L2 repeat: matches 2577. .2691 of consensus"
62253. .62477
/note="MIR repeat: matches 20. .252 of consensus"
62484. .63537
/note="MER10 repeat: matches 1. .1071 of consensus"
63966. .65052
/note="L1MB8 repeat: matches 5088. .6173 of consensus"
68422. .69775
/note="Charliela repeat: matches 7. .1455 of consensus"
```

```
repeat_region 70089. .70150
/note="31 copies 2 mer at 79% conserved"
repeat_region 70091. .70150
/note="15 copies 4 mer atat 80% conserved"
misc-feature 71822. .72082
/note="Sequence from overlapping clone AU356425. Assembly
confirmed by restriction digests."
repeat_region 72162. .72468
/note="AluY repeat: matches 1. .305 of consensus"
72866. .73038
/note="MER5A repeat: matches 7. .189 of consensus"
73289. .73615
/note="AluSx repeat: matches 1. .302 of consensus"
repeat_region 74557. .74623
/note="MER47 repeat: matches 1. .66 of consensus"
repeat_region 74977. .75018
/note="21 copies 2 mer aa 76% conserved"
75398. .75700
/note="AluSx repeat: matches 1. .305 of consensus"
repeat_region 76142. .76605
/note="L1ME2 repeat: matches 4902. .5347 of consensus"
76606. .76911
/note="AluY repeat: matches 1. .312 of consensus"
76912. .77672
/note="L1ME2 repeat: matches 5347. .6155 of consensus"
77727. .77781
/note="MER5A repeat: matches 111. .167 of consensus"
```

```
Query Match 15.1%; Score 37; DB 9; Length 109101;
Best Local Similarity 57.3%; Pred. No. 0.47;
Matches 67; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
```

```
QY 6 ctgggtttccattatatacctctgcatactctctgaagcagcctctct 65
Db 17946 CTCCTTCTCTTATTTCTTCTCTCTCTCTCTCTCTCTCTCTCT 17887
QY 66 ctaagtcacgaagactgcctctactctgtacactgccttactcaagatttt 122
Db 17886 TTCTGCTGAGAGACTGCCCTTCCTGCTTTCGATTTTCACACCTGCATCTCT 17830
```

```
RESULT 9
AC068545 170695 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 6 clone RP11-385M12, WORKING DRAFT
DEFINITION AC068545
ACCESSION AC068545.3 GI:8099098
VERSION AC068545.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 170695)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL unpublished
REFERENCE 2 (bases 1 to 170695)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On May 26, 2000 this sequence version replaced gi:7820143.
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0385M12
----- Summary Statistics -----
Sequencing vector: M13; 100%
```



```

* 31908 32007: gap of unknown length
* 32008 36364: contig of 4357 bp in length
* 36365 36464: gap of unknown length
* 36465 43309: contig of 6845 bp in length
* 43310 43409: gap of unknown length
* 43410 48185: contig of 4776 bp in length
* 48186 48285: gap of unknown length
* 48286 55626: contig of 7341 bp in length
* 55627 55727: gap of unknown length
* 55727 65781: contig of 10055 bp in length
* 65782 65881: gap of unknown length
* 65882 69673: contig of 3792 bp in length
* 69674 69773: gap of unknown length
* 69774 79389: contig of 9616 bp in length
* 79390 79490: gap of unknown length
* 79490 91453: contig of 11964 bp in length
* 91454 91553: gap of unknown length
* 91554 96671: contig of 5118 bp in length
* 96672 96771: gap of unknown length
* 96772 100544: contig of 3773 bp in length
* 100545 100644: gap of unknown length
* 100645 111423: contig of 10779 bp in length
* 111424 111523: gap of unknown length
* 111524 115278: contig of 3755 bp in length
* 115279 115378: gap of unknown length
* 115379 124011: contig of 8633 bp in length
* 124012 124111: gap of unknown length
* 124112 138475: contig of 14364 bp in length
* 138476 138575: gap of unknown length
* 138576 140613: contig of 2038 bp in length
* 140614 140713: gap of unknown length
* 140714 144949: contig of 4236 bp in length
* 144950 145049: gap of unknown length
* 145050 164352: contig of 19303 bp in length
* 164353 164452: gap of unknown length
* 164453 170540: contig of 6088 bp in length
* 170541 170640: gap of unknown length
* 170641 185046: contig of 14406 bp in length
* 185047 185146: gap of unknown length
* 185147 222309: contig of 37063 bp in length
* 222310 222309: gap of unknown length
* 223217 253217: contig of 30908 bp in length.

```

FEATURES

```

source
    1..253217
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="19"
        /clone_id="CTD-3220F14"
        /clone_lib="Caltech human BAC library D"
BASE COUNT      62246 a 62867 c 62532 g 61752 t 3820 others
ORIGIN

```

```

Query Match      15.1% Score 37; DB 2; Length 253217;
Best Local Similarity 64.7% Pred. No. 0.5;
Matches 55; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
QY 11 ttcttcattatctatctcgcattactctctcgcagtcagagccttctcttaag 70
Db 187609 ttttctctcattatgattcctcttatttatttattatgagacgaattcgcctctgtctcac 187550
QY 71 tcacgggaactgcccttgactactg 95
Db 187549 acgagagatcgccgacatgccatgtg 187525

```

```

RESULT 11
AC069411 156733 bp DNA linear PRI 27-SEP-2001
LOCUS Homo sapiens 3 BAC RP11-609K13 (Roswell Park Cancer Institute Human
DEFINITION BAC library) complete sequence.
ACCESSION AC069411
VERSION AC069411.21 GI:13876417

```

KEYWORDS

HTG.
human.

SOURCE
Homo sapiens

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS
1 (bases 1 to 156733)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-oshan,F.R., Allen,C.,
Alshrocks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Burich,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dim,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earmhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A.,
He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M.,
Holloway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Huik,S.,
Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loui-seged,H., Lorado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newton,J., Newton,N., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwokenkwo,S., Oguh,M., Okunou,G., Oragunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S.,
Scott,G., Shen,H., Shoohtari,N., Sison,I., Sodergren,E.,
Sonlike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svarek,A.,
Tabors,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
Taylor,T., Tellrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Wallington,C., Williams,G.,
Williamson,A., Wleczek,R., Woodson,S., Worley,K., Wu,C.,
Wu,Y.F., Zhou,J., Zorilla,S., Naylor,S.L. and Gibbs,R.

TITLE

JOURNAL
Unpublished

REFERENCE

AUTHORS
Worley,K.C.
JOURNAL
Submitted (30-MAY-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS
Worley,K.C.
JOURNAL
Submitted (28-APR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS
Worley,K.C.
JOURNAL
Submitted (01-MAY-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS
Worley,K.C.
JOURNAL
Submitted (27-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS
Worley,K.C.
JOURNAL
Submitted (27-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS
Worley,K.C.
JOURNAL
Submitted (27-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS
Worley,K.C.
JOURNAL
Submitted (27-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS
Worley,K.C.
JOURNAL
Submitted (27-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Apr 28, 2001 this sequence version replaced gi:13811719.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-helpebcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES
Location/Qualifiers
source

```
1. 156733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-609K13"
repeat_region
1. 147
/rpt_family="L1M2"
repeat_region
218. 428
/rpt_family="L1M2"
repeat_region
691. 1451
/rpt_family="L1M2"
repeat_region
1481. 3870
/rpt_family="L1MA4A"
repeat_region
5874. 6212
/rpt_family="L1M1"
repeat_region
6209. 6520
/rpt_family="L1MA4"
repeat_region
6546. 6580
/rpt_family="AT_rich"
repeat_region
7256. 7277
/rpt_family="AT_rich"
repeat_region
7543. 7626
/rpt_family="MIR"
repeat_region
8402. 8442
/rpt_family="GA-rich"
repeat_region
complement(8762..8985)
/rpt_family="MER20"
repeat_region
complement(10497..10671)
/rpt_family="MER117"
STS
10795..11074
/standard_name="G54755"
/db_xref="dbSTS:83999"
repeat_region
11455..11758
/rpt_family="AluJo"
repeat_region
12083..12104
```

```
repeat_region /rpt_family="AT_rich"
14223..14244
repeat_region /rpt_family="AT_rich"
14275..14357
repeat_region /rpt_family="CT-rich"
complement(14363..14441)
/rpt_family="L2"
repeat_region 17006..17091
/rpt_family="CT-rich"
17100..17190
/rpt_family="(TTCC)n"
17199..17237
/rpt_family="T-rich"
repeat_region complement(17238..17402)
/rpt_family="AluJ"
17403..17419
/rpt_family="T-rich"
repeat_region complement(17420..23556)
/rpt_family="L1PA3"
17755..17886
/standard_name="G19948"
/db_xref="dbSTS:32826"
complement(23564..23691)
/rpt_family="AluJ"
23694..23796
/rpt_family="L2"
repeat_region complement(23979..24182)
/rpt_family="MER20"
repeat_region complement(24300..24457)
/rpt_family="L2"
repeat_region complement(24541..25128)
/rpt_family="L2"
repeat_region 26769..27238
/rpt_family="MER74A"
27240..27749
/rpt_family="MLT2B"
27750..27830
/rpt_family="MER74A"
repeat_region complement(28216..28522)
/rpt_family="Tigger4(Comb1)"
repeat_region complement(28538..29878)
/rpt_family="Tigger4(Comb1)"
repeat_region complement(29873..30262)
/rpt_family="Tigger4(Comb1)"
30219..30269
/rpt_family="MER46A"
repeat_region 31408..31531
/rpt_family="MER5B"
repeat_region 31534..31570
```

Query Match 15.0%; Score 36.8; DB 9; Length 156733;
Best Local Similarity 49.0%; Pred. No. 0.56;
Matches 98; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

```
QY 2 ttaacggagtttccattatcatcctcgcattactctcctgaagtcgctctt 61
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34147 TTATATGGCTTTGACATTTGTAAGTATGTTGTTCCCTGATGGAATGATCT 34206
62 ctctcctaagtcacgggaacgcctcctgctactcgtgaccccttactcagcagttt 121
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34207 TTATTTGGCGTATATCTCTGAGTTCTTGCTCTCTCTCACTTCGGCATTTGCTATGCAATTTAT 34266
122 tgtctcgggaacccctggaattctgctaataactataactcgtagtgctgaaggaaaca 181
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34267 ATGAAATGGCACAAGCTTGTCAGTCAATCACTAGAGGTGTAAGTGGACATGCGCTCCT 34326
182 gatgaagaacatgacctca 201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 34327 GCTTAAAGCCTTAAATACAA 34346
```

RESULT 12
AC073093

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 1, 2001 this sequence version replaced gi:14626346.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L2164

Center clone name: 222_N13

FEATURES

```

source
    Location/Qualifiers
      1..192696
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /map="11"
        /clone="RP11-222N13"
        /clone_lib="RPC1-11 Human Male BAC"
        complement(275..339)
        /rpt_family="MER2"
        complement(392..640)
        /rpt_family="MER2"
        complement(701..891)
        /rpt_family="MIR"
        965..1054
        /rpt_family="MIR"
        1092..1269
        /rpt_family="MIR"
        complement(2758..3015)
        /rpt_family="L3"
        complement(3071..6344)
        /rpt_family="L2"
        complement(6669..6848)
        /rpt_family="MIR"
        7341..7527
        /rpt_family="MIR"
        8125..8262
        /rpt_family="MIR"
        8450..8714
        /rpt_family="LHIS"
        8715..8758
        /rpt_family="A-rich"
        10475..10506
        /rpt_family="(TRG)n"
        12602..12671
        /rpt_family="CT-rich"
        14010..14030
        /rpt_family="(A)n"
        14046..14104
        /rpt_family="GA-rich"
        17600..17802
        /rpt_family="MIR"
        18161..18526
        /rpt_family="L2"
        19121..19164
        /rpt_family="MERSA"
        complement(19309..19393)
        /rpt_family="MIR"
        20228..20329
        /rpt_family="L2"
        20431..20466
        /rpt_family="MIR"
        complement(20863..20943)
        /rpt_family="MIR"
        complement(21022..21558)
        /rpt_family="MER106B"
        21270..21565
        /rpt_family="Alusp"
        21487..21493
        unsure

```

```

unsure /note="<30 qual SNGL region"
21489 /note="probably G"
repeat_region complement(21650..22214)
unsure /rpt_family="MULTIG1"
22001..22007 /note="<30 qual SNGL region"
unsure 22337..22362 /note="<30 qual SNGL region"
repeat_region complement(22660..22875)
23270..23332 /rpt_family="Alusg/x"
repeat_region /rpt_family="AT-rich"
repeat_region complement(24498..24690)
repeat_region /rpt_family="MLT1K"
/rpt_family="MLT1K" complement(24717..24959)
25186..25310 /rpt_family="MLT1K"
repeat_region 25186..25310
/rpt_family="L2"
repeat_region 25588..25745
/rpt_family="MERS8A"
repeat_region 25746..26132
/rpt_family="MSRA"
26133..26180 /rpt_family="MERS8A"
repeat_region /rpt_family="MERS8A"
26188..26222 /rpt_family="AT-rich"
repeat_region complement(26542..26735)
/rpt_family="MERS8B"
26821..26971
/rpt_family="MIR"
repeat_region complement(26984..27583)
/rpt_family="L2"
repeat_region complement(27729..27896)
/rpt_family="HAL1"
repeat_region complement(28051..28407)
/rpt_family="TRHB1B"
repeat_region complement(28493..28900)
/rpt_family="HAL1"
28901..29330 /rpt_family="MSRB"
repeat_region complement(29331..29807)
/rpt_family="HAL1"
repeat_region complement(29875..29992)
/rpt_family="L2"
30053..30088 /rpt_family="(TRCA)n"
repeat_region complement(30089..30195)
/rpt_family="L2"
repeat_region complement(30207..30999)
/rpt_family="HAL1"
repeat_region 31480..31538
/rpt_family="L2"
32597..32694 /rpt_family="MIR"
repeat_region 32695..32855
/rpt_family="MERSA"
repeat_region complement(33143..33204)
/rpt_family="L2"
33799..33985 /rpt_family="MERS3"
repeat_region 35105..35269
/rpt_family="MER117"
repeat_region 35475..35905
/rpt_family="LIM4"
35914..36300 /rpt_family="LIM4"
repeat_region complement(36314..36411)
/rpt_family="L2"
repeat_region complement(36663..36777)

```

Query Match 14.4%; Score 35.4; DB 9; Length 192696;
 Best Local Similarity 52.3%; Pred. No. 1.7;

	Matches	78;	Conservative	0;	Mismatches	71;	Indels	0;	Gaps	0;
Oy	42	tcctgagtcagagccctctctctcctaagtcagggaaactgccccttgctacttgacct	101							
Db	51904	tctcttattcacagcattgacattttccctcagagaaactctgccttttacctgtgaatt	51963							
Oy	102	gcccttactcagcagtttctgctggaagccctggatctgctaataatactacact	161							
Db	51964	tcagtttgctgaaacttttttttttttttgccacacattggccatggaatttctcttact	52023							
Oy	162	gtagtgctgaaaggaacagatgaaga	190							
Db	52024	ttctaggagagccaggaatgcatttaagaa	52052							

Search completed: September 26, 2002, 04:22:05
Job time: 8890 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 01:52:05 ; Search time 475.38 Seconds
(without alignments)
884.859 Million cell updates/sec

Title:	US-09-874-390-1_COPY_1_245
Perfect score:	245
Sequence:	1 attactggtttctatt.....caagctgacgctgcaag 245

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

```
Searched:      1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 34728722
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database : N_Genseq_032802 : *

1:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1980.DAT *
2:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1981.DAT *
3:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1982.DAT *
4:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1983.DAT *
5:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1984.DAT *
6:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1985.DAT *
7:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1986.DAT *
8:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1987.DAT *
9:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1988.DAT *
10:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1989.DAT *
11:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1990.DAT *
12:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1991.DAT *
13:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1992.DAT *
14:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1993.DAT *
15:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1994.DAT *
16:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1995.DAT *
17:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1996.DAT *
18:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1997.DAT *
19:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1998.DAT *
20:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA1999.DAT *
21:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA2000.DAT *
22:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA2001A.DAT *
23:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA2001B.DAT *
24:	/SIDS1/gcgdata/genseq/genseqn-emb1/NA2002.DAT *

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	245	100.0	2319	21	AAHA6623	Human C2/4GNT CDNA
2	182	74.3	2236	22	AAH34463	Human colon cancer
3	176	71.8	2229	21	AAAC99109	Human pancreatic c
4	133	54.3	210	21	AAAC98742	Human colon cancer
5	31.8	13.0	337	22	AAAS03444	DNA encoding anti
6	31.8	13.0	330	22	AAAS03458	DNA encoding anti
7	31.4	12.8	2965	22	AAH48118	Dihydroroctase 9 c
8	31	12.7	6530	22	AAH51101	Human Immune Themia
9	30.8	12.6	319608	21	AAH51601	Human chromosome 1

C	10	30.8	12.6	319608	22	AAS03031	Human schistocerca
C	11	30.4	12.4	2464	22	AAK72739	Human immune/haema
C	12	30.4	12.4	3425	22	AAK72740	Human immune/haema
C	13	30.4	12.4	3425	22	AAK72741	Human immune/haema
C	14	30.2	12.3	330	22	AAS03512	DNA encoding anti-
C	15	30.2	12.3	330	22	AAS03529	DNA encoding anti-
C	16	30.2	12.3	333	22	AAS03485	DNA encoding anti-
C	17	30.2	12.3	336	22	AAS03451	DNA encoding anti-
C	18	30.2	12.3	360	22	AAS03442	DNA encoding anti-
C	19	30.2	12.3	2124	22	AAK73947	Human immune/haema
C	20	30.2	12.3	2124	22	AAK73948	Human immune/haema
C	21	30.2	12.3	2124	22	AAK73949	Human immune/haema
C	22	29.6	12.1	330	22	AAS03527	DNA encoding anti-
C	23	29.2	11.9	531	22	ABA20908	Human nervous syst
C	24	29.2	11.9	1563	22	AAI70508	Soybean tyrosine G
C	25	29.2	11.9	11092	22	ABA20907	Human nervous syst
C	26	29.2	11.9	11092	22	ABA04509	Human reproductive
C	27	29.2	11.9	106746	21	AAAF210225	Human PCrTA-1 genom
C	28	29.2	11.9	611590	21	AAAF22303	Arabidopsis thailia
C	29	29	11.8	2766	21	AAAC76955	Human OREF ORF2510
C	30	28.8	11.8	400	22	AAI00249	Human reproductive
C	31	28.8	11.8	487	22	AAK82851	Human immune/haema
C	32	28.8	11.8	868	21	AAC32677	Arabidopsis thailia
C	33	28.8	11.8	914	22	AAS27373	CDNA encoding nove
C	34	28.8	11.8	1216	22	AAK58219	Human polynucleoti
C	35	28.8	11.8	1523	22	AAH18378	Human CDNA sequenc
C	36	28.8	11.8	1646	22	AAH14831	Human CDNA sequenc
C	37	28.8	11.8	1861	22	AAHA5803	Human RNA helicase
C	38	28.8	11.8	1899	22	AAH78914	Human sperm-specifi
C	39	28.8	11.8	1949	22	AAK51810	Human polynucleoti
C	40	28.8	11.8	2129	22	AAH34400	Human colon cancer
C	41	28.8	11.8	2283	21	AAAT7749	Human OREF ORF2974
C	42	28.8	11.8	2602	21	AAAT7416	CDNA encoding a hu
C	43	28.8	11.8	4120	22	AAI59717	Human polynucleoti
C	44	28.8	11.8	4161	22	AAI59731	Human polynucleoti
C	45	28.8	11.8	5406	23	ABLO6257	Drosophila melanog

ALIGNMENTS

XX	RESULT	1
AA48623	AAA48623	
ID	AAA48623	standard; cDNA; 2319 BP.
XX	AAA48623;	
AC		
XX		
DT	19-SEP-2000	(first entry)
XX		
DE	Human C2/AGnt	cDNA.
XX		
KW	Human; C2/4Gnt;	UDP-N-acetylglucosamine: O-glycan biosynthesis
KM	O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer;	ss.
XX		
OS	Homo sapiens.	
.XX		
FH	Key	Location/Qualifiers
FT	CDS	496..1812
FT		/*tag= a
FT		/product= "C2/4Gnt"
FT	primer_bind	586..605
FT		/*tag= b
FT	primer_bind	1794..1813
FT		/*tag= c
FT	polyA_signal	2244..2249
FT		/*tag= d
PN	WO200034449-A2.	
PD		
XX	15-JUN-2000.	
PF	03-DEC-1999;	99WO-DK00677.
XX		

PR 04-DEC-1998; 98DK-0001605.
XX (CLAU/) CLAUSEN H.
XX
XX Clausen H, Schwientek T;
PI
XX WPI: 2000-423407/36.
DR P-PSDB: AAY94492.
XX
XX New nucleic acid molecule encoding UDP-N-acetylglucosamine useful as
PT probe for the detection of specified glucosaminyltransferase from
PT other species and related organisms
XX
XX Claim 5; Fig 2; 47pp; English.
XX
XX The present sequence encodes human UDP-N-acetylglucosamine:
CC N-acetylglucosamine beta-1,6-N-acetylglucosaminyltransferase
CC (C2/4Gnt). The protein is the third member of the family of O-glycan
CC beta-1,6-N-acetylglucosaminyltransferases. It adds core 2 or core 4
CC based O-glycans on to oligosaccharides, glycoproteins and
CC glycosphingolipids. C2/4Gnt can therefore be used in the production of
CC appropriately glycosylated glycoconjugates with particular enzymatic,
CC immunogenic, or other biological or physical properties. The nucleotide
CC sequence is useful as a probe for the detection of C2/4Gnt from other
CC species and related organisms and for the recombinant production of
CC C2/4Gnt polypeptide. The nucleotide sequence was identified by analysis
CC of EST database sequence information. Oligonucleotides derived from EST
CC clone 17865 of ATCC were used to isolate two full-length C2/4Gnt clones
CC from a human foreskin genomic PI library by 5' RACE PCR. RT-PCR was
CC performed using Colo205 human cell line mRNA in order to produce cDNA
CC for expression of C2/4Gnt in Sf9. The control of O-glycan core assembly
CC has been implicated in tumour progression and metastasis.
XX
XX Sequence 2319 BP; 615 A; 528 C; 559 G; 617 T; 0 other;
SQ
Query Match 100.0%; Score 245; DB 21; Length 2319;
Best Local Similarity 100.0%; Pred. No. 7, 4e-74;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 attactgggtttccattatcatcctcctgcattacttctctgagtcagacctct 60
Db 1 attactgggtttccattatcatcctcctgcattacttctctgagtcagacctct 60
QY 61 tctcttaagtcaagggaactgacctgctactgtgactgaccttactcagcagttt 120
Db 61 tctcttaagtcaagggaactgacctgctactgtgactgaccttactcagcagttt 120
QY 121 ttgttcgggaagccctgggattctgttaactactcactgtgagtgccgaaggaac 180
Db 121 ttgttcgggaagccctgggattctgttaactactcactgtgagtgccgaaggaac 180
QY 181 agatgaagacatgacctcaagaagctcctgtcaatgagaagcaagctgaagcctgg 240
Db 181 agatgaagacatgacctcaagaagctcctgtcaatgagaagcaagctgaagcctgg 240
QY 241 caaag 245
Db 241 caaag 245
RESULT 2
AAH34463
ID AAH34463 standard; cDNA; 2236 BP.
XX
XX AAH34463;
AC
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:1545.
DE
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 15; ss.

XX
OS Homo sapiens.
XX
XX WO200122920-A2.
XX
XX
XX 05-APR-2001.
PD
XX
XX 28-SEP-2000; 2000MO-US26524.
PF
XX
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI
XX WPI: 2001-235357/24.
DR P-PSDB: AAG75058.
XX
XX Nucleic acids encoding 427 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 1; Page 3190-3191; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 2236 BP; 609 A; 503 C; 548 G; 574 T; 2 other;
SQ
Query Match 74.3%; Score 182; DB 22; Length 2236;
Best Local Similarity 100.0%; Pred. No. 3, 5e-52;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 64 ctctaagtcacgggaactgacctgtgactactgtgacctgaccttactcagcagttttg 123
Db 2 ctctaagtcacgggaactgacctgtgactactgtgacctgaccttactcagcagttttg 123
QY 124 ttctgggaagccctgggattctgttaactactcactcactgtgagtgccgaaggaacaga 183
Db 62 ttctgggaagccctgggattctgttaactactcactcactgtgagtgccgaaggaacaga 121
QY 184 tgaagaacatgacctcaagaagctcctgtcaatgagaagcaagctgaagcctggcctggc 243
Db 122 tgaagaacatgacctcaagaagctcctgtcaatgagaagcaagctgaagcctggcctggc 181
QY 244 ag 245
Db 182 ag 183
RESULT 3
AAC99109
ID AAC99109 standard; cDNA; 2229 BP.
XX
XX AAC99109;
AC

[illegible]

RESULT 4
 ID AAC98742 standard; cDNA. 210 BP.
 AC AAC98742;
 XX AAC98742;
 XX 09-MAR-2001 (first entry)
 XX
 DE Human colon cancer antigen nucleotide sequence SEQ ID NO:752.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antineoplastic; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder; ss.
 XX
 XX Homo sapiens.
 XX
 PN WO200055351-A1.
 PD 21-SEP-2000.
 PF 08-MAR-2000; 2000WO-US05883.
 PR 12-MAR-1999; 99US-0124270.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Ruben SM;
 DR WPI: 2000-587534/55.
 DR P-PSDB: AAB53985.
 XX
 XX Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -
 XX
 PS Claim 1; Page 1310; 2104pp; English.
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB5334 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotropic, antineoplastic and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.
 XX
 XX Sequence 210 BP; 54 A; 47 C; 49 G; 58 T; 2 other;
 Query Match 54.3%; Score 133; DB 21; Length 210;
 Best Local Similarity 98.6%; Pred. No. 9e-36;
 Matches 144; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 64 ctctaaagtcacgggaactgcctctgactactgtgaactgccttactcaagcatttttg 123

Db 53 tcaccattctctctgtgaagaccctccaacatigtggaataatattgtctctgttacc 112
QY 171 gaaggaaacagatgaagacatgaccccaagagcttctctcatgtggaagaccacg 230
Db 113 aacagacaccagcaagcccccgaactcatgtattatgagtcagcaagcgccctcag 172
QY 231 tgacgcctggc 241
Db 173 gggctccctgac 183

RESULT 7

AAH48118
ID AAH48118 standard; cDNA; 2966 BP.
XX
XX
AC AAH48118;
XX
DT 19-SEP-2001 (first entry)
XX
XX DE Dihydroorotase 9 coding sequence.
XX
KW Dihydroorotase 9; virucide; haemostatic; immunomodulatory;
KW antiinflammatory; cytostatic; gene therapy; malignant neoplasm;
KW haemopathy; HIV infection; immunological disease; inflammation; ss.
XX
OS Unidentified.
XX
XX MO200148159-A1.
XX
XX PD 05-JUL-2001.
XX
XX PF 25-DEC-2000; 2000WO-CN00702.
XX
XX PR 27-DEC-1999; 99CN-0125387.
XX
XX (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX
XX PI Mao Y, Xie Y;
XX
XX DR WPI: 2001-418257/44.
XX
XX DR P-PSDB; AAG64239.
XX
XX PT Dihydroorotase 9 and encoded polynucleotide, applicable in diagnosis
XX and treatment of malignant neoplasm, hemopathy, HIV infection,
XX immunological diseases and various inflammation
XX
XX PS Claim 6; Page 27-28; 33pp; Chinese.
XX
XX CC The present sequence is the coding sequence of dihydroorotase 9.
XX CC Dihydroorotase 9 and its coding sequence are useful for the diagnosis and
XX CC treatment of malignant neoplasm, haemopathy, HIV infection, immunological
XX CC diseases and various inflammations.
XX
XX SQ Sequence 2966 BP; 567 A; 786 C; 846 G; 767 T; 0 other;

Query Match 12.8%; Score 31.4; DB 22; Length 2966;
Best Local Similarity 51.0%; Pred. No. 2.7; Mismatches 71; Indels 0; Gaps 0;
Matches 74; Conservative 0;

QY 11 ttctctattatctatctcctgcagcatctctctgcagagcctctctctcctcag 70
Db 658 tgtctcgtattctcttcttgctgctgagtgagggcctgcagctgctctctcgtgtt 717
QY 71 tcaagggaactgccttgcactctgtgacctgccttaccagcagttttgtctggg 130
Db 718 caacaggaactagatcatctacttggtgctgtgtcaccctgggccaatgctcatgt 777
QY 131 aagccctggatctgctataacct 155
Db 778 gaggcctggtatctcagctagacct 802

RESULT 8
AAK61101/c
ID AAK61101 standard; cDNA; 630 BP.
XX
XX AC AAK61101;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:6161.
XX
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157182-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US01354.
XX
XX PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220964.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0228343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.

KW sbg1; g34665; sbg2; g35017; g35018; chromosome 13q31-q33; haplotype;
 KW diallelic marker; polymorphism; schizophrenia; bipolar disorder; ds.
 OS Homo sapiens.
 XX MO200058510-A2.
 XX 05-OCT-2000.
 PD 30-MAR-2000; 2000WO-IB00435.
 PF 30-MAR-1999; 99US-0126903.
 PR 30-APR-1999; 99US-0131971.
 PR 30-APR-1999; 99US-0132065.
 PR 14-JUL-1999; 99US-0143928.
 PR 27-JUL-1999; 99US-0145915.
 PR 29-JUL-1999; 99US-0146452.
 PR 29-JUL-1999; 99US-0146453.
 PR 28-OCT-1999; 99US-0162288.
 XX (GEST) GENSET.
 PA Cohen D, Blumenfeld M, Chumakov I, Bougueleret L, Bihain B;
 PI Essioux L;
 XX WPI: 2000-619082/59.
 DR Polynucleotides comprising sequences from sbg1 and g35018 diallelic
 PT markers are used for genotyping and detecting schizophrenia or bipolar
 PT disorder and predisposition to these disorders -
 PS Claim 1; Page 409-493; 737pp; English.
 XX AAH51601 represents a human genomic nucleotide sequence comprising sbg1,
 CC g34665, sbg2, g35017 and g35018 nucleic acid sequences located on the
 CC human chromosome 13q31-q33 locus. The nucleotide sequences contain
 CC diallelic markers and polymorphisms. Sequences AAH51602 - AAH51626 and
 CC AAB62907 - AAB62915 represent cDNA human sbg1 cDNA sequences and protein
 CC products. AAH51627 - AAH51631 and AAB62916 - AAB62918 represent g35018
 CC cDNA sequences and protein products. Primers AAH51632 - AAH51699 are used
 CC to isolate sbg1 cDNAs, while sbg1 exons from different primates are
 CC represented by sequences AAH51642 - AAH51699. Nucleotide sequences of
 CC amplicons which comprise diallelic markers located on the chromosome
 CC 13q31-q33 locus are represented in AAH51700 - AAH51817. Diallelic markers
 CC are represented in the sequences by degenerate/undefined base codes. PCR
 CC primers AAH51818 and AAH51819 are used in the isolation of sequences of
 CC the invention. The diallelic marker containing nucleotide sequences are
 CC used to determine the identity of the nucleotide at a diallelic marker in
 CC a sample DNA sequence. The nucleotide sequences may be labelled and used
 CC for genotyping by determining the identity of a nucleotide at a Region
 CC D-related diallelic marker in a biological sample from single or multiple
 CC subjects. By determining the frequency of a diallelic marker in a
 CC population an association between a genotype and a trait, a haplotype and
 CC a trait and a phenotype and a trait can be detected. The sequences can be
 CC used to determine a predisposition to or early onset of schizophrenia or
 CC bipolar disorder or a beneficial response to or side effects related to
 CC treatment against schizophrenia or bipolar disorder.
 XX Sequence 319608 BP; 101600 A; 56677 C; 58335 G; 102722 T; 274 other;

Query Match 12.6%; Score 30.8; DB 21; Length 319608;
 Best Local Similarity 54.4%; Pred. No. 36;
 Matches 62; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 11 ttctctattatctactctcgactactctctgagcagcctctctctcaag 70
 Db 66134 TTTTCAGTATGTTCTCTTATCTTTATGTAAACCTCAGTATATGTCGAG 66075
 OY 71 tcacggagacgcctctgactgtgacgccttactcagcagttttgt 124
 Db 66074 TCTCTAAATCTACCCCTCTATCTGTAATGTTCTTTCTATATGCGTGTGT 66021

RESULT 10
 ID AAS09301/C
 XX AAS09301 standard; DNA; 319608 BP.
 AC AAS09301;
 XX 26-SEP-2001 (first entry)
 DE Human schizophrenia associated gene g35030 and diallelic markers A1-A71.
 XX Human; g35030; diallelic marker; A1-A71; chromosome 13q31-q33;
 KW schizophrenia; bipolar disorder; ds.
 OS Homo sapiens.
 XX Key
 FT primer_bind
 FT Location/Qualifiers
 FT 7938..7958
 FT /tag= a
 FT /note= "Binds primer 99-27943..rp"
 FT 8297..8315
 FT /tag= b
 FT /note= "Binds primer 99-27943-150.mis"
 FT 8304..8328
 FT /tag= c
 FT /bound_moiety= Probe_99-27943-150
 FT 8316
 FT /tag= d
 FT /note= "Biallelic marker A1"
 FT complement (8317..8335)
 FT /tag= e
 FT /note= "Binds primer 99-27943-150.mis complement"
 FT complement (8446..8465)
 FT /tag= f
 FT /note= "Binds primer 99-27943..pu complement"
 FT 21365..21385
 FT /tag= g
 FT /note= "Binds primer 99-27935..rp"
 FT 21653..21671
 FT /tag= h
 FT /note= "Binds primer 99-27935-193.mis"
 FT 21660..21684
 FT /tag= i
 FT /bound_moiety= Probe_99-27935-193
 FT 21672
 FT /tag= j
 FT /note= "Biallelic marker A2"
 FT complement (21673..21691)
 FT /tag= k
 FT /note= "Binds primer 99-27935-193.mis complement"
 FT complement (21845..21864)
 FT /tag= l
 FT /note= "Binds primer 99-27935..pu complement"
 FT 65463..65471
 FT /tag= m
 FT /note= "Binds primer 8-128:pu"
 FT 65466..65484
 FT /tag= n
 FT /note= "Binds primer 8-128-33.mis"
 FT 65473..65497
 FT /tag= o
 FT /bound_moiety= Probe_8-128-33
 FT 65485
 FT /tag= p
 FT /note= "Biallelic marker A3"
 FT complement (65486..65504)
 FT /tag= q
 FT /note= "Binds primer 8-128-33.mis complement"
 FT complement (65856..65874)
 FT /tag= r
 FT /note= "Binds primer 8-128..rp complement"
 FT 95034..95053
 FT /tag= s

```
FT primer_bind /note="binds primer 99-31960.pu"
FT 95377..95395
FT /tag- t
FT /note="binds primer 99-31960-363.mis"
FT 95384..95408
FT /tag- u
FT /bound_moiety= Probe_99-31960-363
FT 95396
FT /tag- v
FT /note="Biallelic marker A4"
FT complement (95397..95415)
FT /tag- w
FT /note="binds primer 99-31960-363.mis complement"
FT /tag- x
FT /note="binds primer 99-31960.rp complement"
FT 107022..107040
FT /tag- y
FT /note="binds primer 99-24656.pu"
FT 107262..107280
FT /tag- z
FT /note="binds primer 99-24656-260.mis"
FT 107269..107293
FT /tag- aa
FT /bound_moiety= Probe_99-24656-260
FT 107281
FT /tag- ab
FT /note="Biallelic marker A5"
FT complement (107282..107300)
FT /tag- ac
FT /note="binds primer 99-24656-260.mis complement"
FT complement (107495..107513)
FT /tag- ad
FT /note="binds primer 99-24656.rp complement"
FT 160279..160298
FT /tag- ae
FT /note="binds primer 99-24639.rp"
FT 160621..160639
FT /tag- af
FT /note="binds primer 99-24639-163.mis"
FT 160628..160652
FT /tag- ag
FT /bound_moiety= Probe_99-24639-163
FT 160640
FT /tag- ah
FT /note="Biallelic marker A6"
FT complement (160641..160659)
FT /tag- ai
FT /note="binds primer 99-24639-163.mis complement"
FT 160770..160787
FT /tag- aj
FT /note="binds primer 99-24634.pu"
FT complement (160785..160802)
FT /tag- ak
FT /note="binds primer 99-24639.pu complement"
FT 160857..160875
FT /tag- al
FT /note="binds primer 99-24634-108.mis"
FT 160864..160888
FT /tag- am
FT /bound_moiety= Probe_99-24634-108
FT 160876
FT /tag- an
FT /note="Biallelic marker A7"
FT complement (160877..160895)
FT /tag- ao
FT /note="binds primer 99-24634-108.mis complement"
FT complement (161240..161257)
FT /tag- ap
FT /note="binds primer 99-24634.rp complement"
FT 168813..168830
FT /tag- aq
FT /note="binds primer 99-7652.pu"
FT 168955..168973
FT /tag- ar
FT /note="binds primer 99-7652-162.mis"
FT 168962..168986
FT /tag- as
FT /bound_moiety= Probe_99-7652-162
FT 168974
FT /tag- at
FT /note="Biallelic marker A8"
FT complement (168975..168993)
FT /tag- au
FT /note="binds primer 99-7652-162.mis complement"
FT complement (169331..169351)
FT /tag- av
FT /note="binds primer 99-7652.rp complement"
FT 170666..170686
FT /tag- aw
FT /note="binds primer 99-16100.pu"
FT 170791..170809
FT /tag- ax
FT /note="binds primer 99-16100-147.mis"
FT 170798..170822
FT /tag- ay
FT /bound_moiety= Probe_99-16100-147
FT 170810
FT /tag- az
FT /note="Biallelic marker A9"
FT complement (170811..170829)
FT /tag- ba
FT /note="binds primer 99-16100-147.mis complement"
FT complement (171153..171173)
FT /tag- bb
FT /note="binds primer 99-16100.rp complement"
FT 173065..173085
FT /tag- bc
FT /note="binds primer 99-5862.rp"
FT 173339..173357
FT /tag- bd
FT /note="binds primer 99-5862-167.mis"
FT 173346..173370
FT /tag- be
FT /bound_moiety= Probe_99-5862-167
FT 173358
FT /tag- bf
FT /note="Biallelic marker A10"
FT complement (173359..173377)
FT /tag- bg
FT /note="binds primer 99-5862-167.mis complement"
FT complement (173495..173514)
FT /tag- bh
FT /note="binds primer 99-5862.pu complement"
FT 189753..189771
FT /tag- bi
FT /note="binds primer 99-5919.pu"
FT 189938..189956
FT /tag- bj
FT /note="binds primer 99-5919-215.mis"
FT

Query Match 12.6%; Score 30.8; DB 22; Length 319608;
Best Local Similarity 54.4%; Pred. No. 36;
Matches 62; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 11 ttctcattatcatcctcctgcattactctctgagtcagagcctctctctaag 70
Db 66134 ttttcagatattggttcctccttattgtttattgtaaacctcagtgattatgcgag 66075
OY 71 tcacgggaactgcctctgactctgagcctcttactcagcaggttttgt 124
Db 66074 tctctaaatctaccctcctatctgtaattgtctttcttattgctgtgtct 66021

RESULT 11
AAK72739/c
```

ID AAK72739 standard; DNA; 2464 BP.
 XX
 AC AAK72739;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27551.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01354.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0233397.
 PR 14-SEP-2000; 2000US-0233398.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0233400.
 PR 14-SEP-2000; 2000US-0233401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.

PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0233400.
 PR 14-SEP-2000; 2000US-0233401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX
XX Disclosure; SEQ ID NO 27551; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91971. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 2464 BP; 693 A; 560 C; 662 G; 549 T; 0 other;
XX
XX
XX Query Match 12.4%; Score 30.4; DB 22; Length 2464;
XX Best Local Similarity 50.0%; Pred. No. 5.6;
XX Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
XX
XX QY 34 gcattactctcgtgagcagagcctctctctaagtcaggaactgccttgact 93
XX Db 1184 GCTTTTATCTTTAACTCATATATCTTCCAAATGCACTGTAAGTCTGAGTTCT 1125
XX QY 94 tctgacctgccttactcagcagcttttctctggaagccctggaactctgtaaac 153
XX Db 1124 TGGCACCACCTTCAGATCAAGCTTTTAAATTCACAAAGCAAGCTCGCTTGTGCTAT 1065
XX QY 154 ctatcactgtagtgctggaaggaacagatg 185
XX Db 1064 GTCACCTTCCCTGTGCGACAGGGGACACCATG 1033
XX
XX
XX RESULT 12
XX AAK72740/C
XX ID AAK72740 standard; DNA; 3425 BP.
XX
XX AAK72740;
XX
XX
XX 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27552.
XX DE Human immune/haematopoietic; immune/haematopoietic antigen; cancer;
XX XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KM cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX

OS Homo sapiens.
XX
XX
XX WO200157182-A2.
XX
XX
XX 09-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184654.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226688.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
XX 25-SEP-2000; 2000US-0234997.
XX 25-SEP-2000; 2000US-0234998.
XX 26-SEP-2000; 2000US-0235484.

PR	27-SEP-2000	2000US-023589.4	
PR	27-SEP-2000	2000US-023589.6	
PR	26-SEP-2000	2000US-023632.7	
PR	26-SEP-2000	2000US-023636.7	
PR	26-SEP-2000	2000US-023636.8	
PR	29-SEP-2000	2000US-023636.9	
PR	29-SEP-2000	2000US-023637.0	
PR	26-SEP-2000	2000US-023637.9	
PR	02-OCT-2000	2000US-023680.2	
PR	02-OCT-2000	2000US-023703.7	
PR	02-OCT-2000	2000US-023703.8	
PR	02-OCT-2000	2000US-023703.9	
PR	02-OCT-2000	2000US-023704.0	
PR	13-OCT-2000	2000US-023939.5	
PR	13-OCT-2000	2000US-023939.7	
PR	20-OCT-2000	2000US-024096.0	
PR	20-OCT-2000	2000US-024112.1	
PR	20-OCT-2000	2000US-024117.85	
PR	20-OCT-2000	2000US-024117.86	
PR	20-OCT-2000	2000US-024117.87	
PR	20-OCT-2000	2000US-024118.08	
PR	20-OCT-2000	2000US-024118.09	
PR	20-OCT-2000	2000US-024118.26	
PR	01-NOV-2000	2000US-024645.17	
PR	01-NOV-2000	2000US-024645.17	
PR	08-NOV-2000	2000US-024647.4	
PR	08-NOV-2000	2000US-024647.5	
PR	08-NOV-2000	2000US-024647.6	
PR	08-NOV-2000	2000US-024647.7	
PR	08-NOV-2000	2000US-024647.8	
PR	08-NOV-2000	2000US-024652.3	
PR	08-NOV-2000	2000US-024652.4	
PR	08-NOV-2000	2000US-024652.4	
PR	08-NOV-2000	2000US-024652.6	
PR	08-NOV-2000	2000US-024652.7	
PR	08-NOV-2000	2000US-024653.2	
PR	08-NOV-2000	2000US-024653.2	
PR	08-NOV-2000	2000US-024656.09	
PR	08-NOV-2000	2000US-024661.0	
PR	08-NOV-2000	2000US-024661.1	
PR	08-NOV-2000	2000US-024661.3	
PR	17-NOV-2000	2000US-024920.7	
PR	17-NOV-2000	2000US-024920.8	
PR	17-NOV-2000	2000US-024920.9	
PR	17-NOV-2000	2000US-024920.9	
PR	17-NOV-2000	2000US-024921.0	
PR	17-NOV-2000	2000US-024921.0	
PR	17-NOV-2000	2000US-024921.6	
PR	17-NOV-2000	2000US-024921.7	
PR	17-NOV-2000	2000US-024921.8	
PR	17-NOV-2000	2000US-024924.4	
PR	17-NOV-2000	2000US-024924.5	
PR	17-NOV-2000	2000US-024924.5	
PR	17-NOV-2000	2000US-024926.4	
PR	17-NOV-2000	2000US-024926.5	
PR	17-NOV-2000	2000US-024929.7	
PR	17-NOV-2000	2000US-024929.8	
PR	17-NOV-2000	2000US-024930.0	
PR	01-DEC-2000	2000US-025016.0	
PR	01-DEC-2000	2000US-025039.1	
PR	05-DEC-2000	2000US-025103.0	
PR	05-DEC-2000	2000US-025119.88	
PR	05-DEC-2000	2000US-025671.9	
PR	06-DEC-2000	2000US-025147.9	
PR	08-DEC-2000	2000US-025185.6	
PR	08-DEC-2000	2000US-025188.8	
PR	08-DEC-2000	2000US-025189.9	
PR	08-DEC-2000	2000US-025196.9	
PR	08-DEC-2000	2000US-025199.0	
PR	11-DEC-2000	2000US-025409.7	
PR	05-JAN-2001	2001US-025967.8	
XX	(HUMA-) HUMAN GENOME SCI INC		

XX	Rosen CA,	Barash SC,	Ruben SW;
PI			
XX	WP1: 2001-483426/52.		
DR			
XX			
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,		
PT	useful for preventing, diagnosing and/or treating cancers and		
PR	metastasis -		
XX			
PS	Disclosure; SEQ ID NO 27552; 3071pp + Sequence Listing; English.		
XX			
CC	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)		
CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic		
CC	activity, and can be used in gene therapy and vaccine production. (I)		
CC	proteins and polynucleotides may be used in the prevention, diagnosis and		
CC	treatment of diseases associated with inappropriate (I) expression. For		
CC	example, they may be used to treat disorders associated with decreased		
CC	expression by rectifying mutations or deletions in a patient's genome		
CC	that affect the activity of (I) by expressing inactive proteins or to		
CC	supplement the patients own production of (I). Additionally, (I)		
CC	polynucleotides may be used to produce the secreted (I), by inserting the		
CC	nucleic acids into a host cell and culturing the cell to express the		
CC	protein. (II) proteins and polynucleotides may be used to prevent,		
CC	diagnose and treat immune/hematopoietic-related diseases, especially		
CC	cancers and cancer metastases of hematopoietic-derived cells. AAK64703		
CC	to AAK7654 represent human immune/hematopoietic antigen genomic		
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169		
XX	represent sequences used in the exemplification of the present invention.		
SQ	Sequence 3425 BP: 916 A; 794 C; 951 G; 764 T; 0 other:		
Query Match	12.4%;	Score 30.4;	DB 22; Length 3425;
Best Local Similarity	50.0%;	Pred. No. 6.5;	
Matches	76; Conservative	0; Mismatches	76; Indels 0; Gaps 0;
OY	34 gcaattactctctcgtagtcagagccctctctctctaagtacgggaatgccttgactact 93		
Db	1184 GCTTTATTCCTTAAACCCATCATATATTCTTCCAATGCAGCTGAAGTGCTGACTTCT 1125		
OY	94 tgtgaccggccctttacctacagcagcttttgtctctcggaagccctggagcttcgctaatac 153		
Db	1124 TGCCACCACCACTTCGAGATCAGAAGCTTTTAAAATTCCAAAGCCAAAGCGCTGCTTCTCTAAT 1065		
OY	154 ctatacatgtagtgctgtaagggaacaagatg 185		
Db	1064 GTACACTTCCCTGTGGCAACGGGCAACCAATG 1033		
RESULT 13			
ID	AAK72741/C		
XX	AAK72741 standard; DNA; 3425 BP.		
AC	AAK72741;		
XX			
DT	06-NOV-2001 (first entry)		
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27553.		
XX			
KW	Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;		
XX	cytostatic; gene therapy; vaccine; metastasis; ds.		
OS	Homo sapiens.		
XX			
WO	WO200157182-A2.		
EN			
XX			
PD	09-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US01354.		
XX			
RR	31-JAN-2000; 2000US-0179065.		
RR	04-FEB-2000; 2000US-0180628.		
RR	24-FEB-2000; 2000US-0184664.		
RR			

PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216680.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI, 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
PS Disclosure; SEQ ID NO 27553; 3071pp + Sequence Listing; English.
XX

CC AAK5491 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cyclostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 3425 BP; 916 A; 794 C; 951 G; 764 T; 0 other;

Query Match 12.4%; Score 30.4; DB 22; Length 3425;
Best Local Similarity 50.0%; Pred. No. 6.5; Matches 76; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 34 gattactctctgagtcagagcctctctctcttaagtcagcggaactgccttgact 93
DB 1184 GCTTTATATCTTTAAACCATATATATCTTCCAAATGACAGTGAAGTCTCTGAGTTCT 1125
QY 94 tctgacctgccttactactacagatttctgtctgggaagccctggattctgctaatac 153
DB 1124 TCCACACCCCTTTCAGATCAAGCTTTTAAATTCACAAAGCCAGCTTCGCTTCTAT 1065
QY 154 ctatactgtagtgctggaaggaagaatg 185
DB 1064 CTCACCTTCCCTGTGTGGCAAGGGCACACATG 1033

RESULT 14

AAS03512
ID AAS03512 standard; cDNA; 330 BP.

XX AAS03512;

DT 29-AUG-2001 (first entry)

DE DNA encoding anti-adipocyte monoclonal antibody light chain, FAT 99.

XX Antibody; adipocyte; heavy chain; light chain; obesity; FAT;
KW heart disease; complementarity determining region; CDR; ss.
XX

OS Homo sapiens.

PN WO200127279-A1.

PD 19-APR-2001.

PF 11-OCT-2000; 2000WO-GB03900.

PR 12-OCT-1999; 99US-0158812.

PA (CAMP-) CAMBRIDGE ANTI-BODY TECHNOLOGY.

PI Edwards BM, Main SH, Vaughan TJ;

DR WPI; 2001-282031/29.

DR P-PSDB; AAU02612.

PT Panel of specific binding members of antibody molecules which bind to
PT whole adipocytes is used in the treatment of obesity and obesity
PT related diseases -
XX
XX

PS Disclosure; Page 162; 182pp; English.

XX AAS03401-AAS03535 represent anti-adipocyte monoclonal antibody heavy
CC and light chain coding sequences of the invention. The antibodies can be
CC used in the treatment of obesity and obesity related diseases. The fat
CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
CC mass of an obese patient or the antibody can be used as a therapeutic
CC itself. Antibodies binding specifically to adipocytes can be used to
CC activate the immune system to destroy the cells by complement mediated
CC lysis. The antibodies may be labeled with a detectable label such as
CC radioisotope, fluorescent or chemical group and used in methods of
CC diagnosis in human subjects e.g. to determine the presence of adipocyte
CC antigen on the surface of an adipocyte to detect or determine the
CC presence or level of adipocytes in a cell or tissue sample. The
CC antibodies can be used as an alternative means of treatment for obese
CC patients other than undergoing surgery to remove excess fat. Antibodies
CC for different types of fat deposits can also be produced e.g. intra-
CC abdominal fat associated with heart disease.
XX
XX
SQ Sequence 330 BP; 75 A; 96 C; 83 G; 76 T; 0 other;

Query Match 12.3%; Score 30.2; DB 22; Length 330;
Best Local Similarity 51.9%; Pred. No. 2.7; Matches 68; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 111 tcagcagtttctgtctgggaagccctgggattctgctaatacctaactgagtgct 170
DB 53 tcacatttcctgcctctggaagcactccaacatttggaataattatgtctctgtgtac 112
QY 171 gaagggaacagatgaagaacatgacctcaaggagcttctctgtaagtgaagaccaga 230
DB 113 aacagacaccagcaagaagccccaactcatgatgatgagtaagcgcctcag 172
QY 231 tgacgcttgc 241
DB 173 gggctccctgac 183

RESULT 15

AAS03529
ID AAS03529 standard; cDNA; 330 BP.

XX AAS03529;

DT 29-AUG-2001 (first entry)

DE DNA encoding anti-adipocyte monoclonal antibody light chain, FAT 113.

XX Antibody; adipocyte; heavy chain; light chain; obesity; FAT;
KW heart disease; complementarity determining region; CDR; ss.
XX

OS Homo sapiens.

PN WO200127279-A1.

PD 19-APR-2001.

PF 11-OCT-2000; 2000WO-GB03900.

PR 12-OCT-1999; 99US-0158812.

PA (CAMP-) CAMBRIDGE ANTI-BODY TECHNOLOGY.

PI Edwards BM, Main SH, Vaughan TJ;

DR WPI; 2001-282031/29.

DR P-PSDB; AAU02629.

PT Panel of specific binding members of antibody molecules which bind to
PT whole adipocytes is used in the treatment of obesity and obesity
PT related diseases -
XX
XX

PS Disclosure; Page 173; 182pp; English.

XX AAS03401-AAS03535 represent anti-adipocyte monoclonal antibody heavy
 CC and light chain coding sequences of the invention. The antibodies can be
 CC used in the treatment of obesity and obesity related diseases. The
 CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
 CC mass of an obese patient or the antibody can be used as a therapeutic
 CC itself. Antibodies binding specifically to adipocytes can be used to
 CC activate the immune system to destroy the cells by complement mediated
 CC lysis. The antibodies may be labeled with a detectable label such as
 CC radiolabel, fluorescent or chemical group and used in methods of
 CC diagnosis in human subjects e.g. to determine the presence of adipocyte
 CC antigen on the surface of an adipocyte to detect or determine the
 CC presence or level of adipocytes in a cell or tissue sample. The
 CC antibodies can be used as an alternative means of treatment for obese
 CC patients other than undergoing surgery to remove excess fat. Antibodies
 CC for different types of fat deposits can also be produced e.g. intra-
 CC abdominal fat associated with heart disease.
 XX
 SQ Sequence 330 BP; 76 A; 96 C; 82 G; 76 T; 0 other;

Query Match 12.3%; Score 30.2; DB 22; Length 330;
 Best Local Similarity 51.9%; Pred. No. 2.7;
 Matches 68; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
 QY 111 tcagcagttttgtctgggaagccctgggattctgctaactatcactgtagtgct 170
 Db 53 tcaccatttcctgctcgtgaagcaccctcccaacttgggaataatattatgctcctggtacc 112
 QY 171 gaagggaacagatgaacatgacctcaaggagcttcctgtcaatgagaagaaccaagc 230
 Db 113 aacagcacccagcgaagaagccccaactcatgattatgattagtaagcgccctcag 172
 QY 231 tgacgcctgagc 241
 Db 173 aggtccctgac 183

Search completed: September 26, 2002, 03:03:33
 Job time: 4288 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 01:54:30 : Search time 94.91 Seconds
(without alignments)
634.077 Million cell updates/sec

Title: US-09-874-390-1_COPY_1_245

Perfect score: 245

Sequence: 1 attaacgtggtttctatt.....caagctgacgcctgcaaa 245

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCtUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32.4	13.2	7218	1	US-08-232-463-14 Sequence 14, Appl
2	29.2	11.9	1563	4	US-09-377-557-11 Sequence 11, Appl
3	28.8	11.8	72604	4	US-09-268-992-7 Sequence 7, Appl
4	28.6	11.7	333	2	US-08-652-816A-27 Sequence 27, Appl
5	28.4	11.6	1907	3	US-08-826-611-1 Sequence 1, Appl
6	28.2	11.5	2174	1	US-07-817-917-2 Sequence 2, Appl
7	27.8	11.3	1542	1	US-08-207-904-1 Sequence 1, Appl
8	27.8	11.3	3706	1	US-08-207-904-16 Sequence 16, Appl
9	27.2	11.1	347	4	US-08-952-061-3 Sequence 3, Appl
10	27.2	11.1	1571	4	US-08-868-435-32 Sequence 32, Appl
11	27.2	11.1	1571	4	US-08-744-231-32 Sequence 32, Appl
12	27	11.0	587	4	US-09-091-590A-2 Sequence 2, Appl
13	27	11.0	1280	4	US-09-276-531-52 Sequence 52, Appl
14	27	11.0	1441	1	US-08-136-277-18 Sequence 18, Appl
15	27	11.0	1441	3	US-08-479-403-18 Sequence 18, Appl
16	27	11.0	1441	3	US-08-835-734-18 Sequence 18, Appl
17	27	11.0	2750	1	US-08-136-277-1 Sequence 1, Appl
18	27	11.0	2750	2	US-08-479-403-1 Sequence 1, Appl
19	27	11.0	2750	2	US-08-835-734-1 Sequence 1, Appl
20	26.8	10.9	3066	1	US-08-142-439A-1 Sequence 1, Appl
21	26.8	10.9	3066	3	US-08-863-790-27 Sequence 27, Appl
22	26.8	10.9	4533	3	US-08-863-790-27 Sequence 27, Appl
23	26.8	10.9	4533	3	US-08-296-749-27 Sequence 27, Appl
24	26.4	10.8	158	4	US-09-019-095A-40 Sequence 40, Appl
25	26.4	10.8	160	4	US-09-019-095A-39 Sequence 39, Appl
26	26.4	10.8	918	3	US-08-937-271-9 Sequence 9, Appl
27	26.4	10.8	1650	4	US-08-868-373-7 Sequence 7, Appl

28	26.4	10.8	1873	4	US-09-019-095A-37 Sequence 37, Appl
29	26.4	10.8	2277	1	US-08-676-967-2 Sequence 2, Appl
30	26.4	10.8	2277	1	US-08-676-974-2 Sequence 2, Appl
31	26.4	10.8	2277	2	US-09-098-487-2 Sequence 2, Appl
32	26.4	10.8	2674	4	US-09-019-095A-1 Sequence 1, Appl
33	26.4	10.8	8700	2	US-08-392-625-16 Sequence 16, Appl
34	26.4	10.8	8700	2	US-08-466-961A-16 Sequence 16, Appl
35	26.4	10.8	8700	2	US-08-645-193B-18 Sequence 18, Appl
36	26	10.6	459	2	US-08-852-807-19 Sequence 19, Appl
37	26	10.6	2107	4	US-09-180-852-1 Sequence 1, Appl
38	26	10.6	2368	1	US-07-698-926A-1 Sequence 1, Appl
39	26	10.6	246240	2	US-08-724-394A-20 Sequence 20, Appl
40	26	10.6	246240	2	US-08-724-394A-21 Sequence 21, Appl
41	26	10.6	246240	2	US-08-724-394A-22 Sequence 22, Appl
42	25.8	10.5	1178	2	US-08-107-676-26 Sequence 26, Appl
43	25.8	10.5	1211	2	US-08-107-676-26 Sequence 2, Appl
44	25.8	10.5	2100	3	US-08-938-830-2 Sequence 2, Appl
45	25.8	10.5	2100	3	US-09-020-222-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgpt-F1s
US-08-232-463-14

[illegible]

```

RESULT 2
US-09-377-557-11/c
; Sequence 11, Application US/09377557
; Patent No. 6297055
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Farnodu, Layo O.
; APPLICANT: Orozco, Emil M. Jr.
; TITLE OF INVENTION: Amino Acid Decarboxylases
; FILE REFERENCE: BB-1237
; CURRENT APPLICATION NUMBER: US/09/377,557
; CURRENT FILING DATE: 1999-08-19
; EARLIER APPLICATION NUMBER: 60/099,493
; EARLIER FILING DATE: September 8, 1998
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 11
; LENGTH: 1563
; TYPE: DNA
; ORGANISM: Glycine max
US-09-377-557-11

```

Query Match	11.9%	Score 29.2	DB 4	Length 1563
Best Local Similarity	65.2%	Pred. No. 1		
Matches 43	Conservative 0	Mismatches 23	Indels 0	Gaps 0
OY	137	tgggattctgcgaatacctacatcagctgtagtctctaagggaacacagatataagaactgac	196	
Db	1207	TTGGACCCCTGGCATTACCAATCTTGCAACTGCAGAAAGTAAAGCTGTAGCAAAACCCAC	1148	
OY	197	ctcaag	202	
Db	1147	AGCAAG	1142	

RESULT 3
 US-09-268-992-7
 : Sequence 7, Application US/09268992
 : Patent No. 6342351
 : GENERAL INFORMATION:
 : APPLICANT: Chen, H.
 : APPLICANT: Freilmer, N.
 : TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
 : TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
 : FILE REFERENCE: 7853-138
 : CURRENT APPLICATION NUMBER: US/09/268,992
 : CURRENT FILING DATE: 1999-03-16
 : EARLIER APPLICATION NUMBER: 09/236,134
 : EARLIER FILING DATE: 1999-01-22
 : EARLIER APPLICATION NUMBER: 60/106,056
 : EARLIER FILING DATE: 1998-10-28
 : EARLIER APPLICATION NUMBER: 60/088,312
 : EARLIER FILING DATE: 1998-06-05
 : EARLIER APPLICATION NUMBER: 60/078,044
 : EARLIER FILING DATE: 1998-03-16

```

: NUMBER OF SEQ ID NOS: 84
: SOFTWARE: FastSeq for Windows Version 3.0.0
: SEQ ID NO: 7
: LENGTH: 72604
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: modified_base
: LOCATION: all n positions
: OTHER INFORMATION: n=a, c, g, or t
US-09-266-992-7

```

Query Match	11.8%	Score 28.8;	DB 4;	length 72604;
Best Local Similarity	62.5%	Pred. No. 9;		
Matches 45; Conservative	0;	Mismatches 27;	Indels 0;	Gaps 0;

Accession	Sequence	Position
Oy 158	catcttaagtcgttgaaaggaacagatgaagacactgacctcaagagcttccttcaat	217
Db 64714	ctctgaagagagaatctggaatttgacacagtcctctgtccccaagagcttcctatca	64773
Oy 218	gagaagaccacaag	229
Db 64774	cagggaccacaag	64785

RESULT 4
US-08-652-816A-27
Sequence 27, Application US/08652816A
Patent No. 5872215
GENERAL INFORMATION:
APPLICANT: Osbourn, JK
APPLICANT: Allen, DJ
APPLICANT: McCafferty, JG
TITLE OF INVENTION: Specific binding members, materials and
TITLE OF INVENTION: methods.
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/652,816A
FILING DATE: 23-MAY-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9125579.4
FILING DATE: 02-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9125579.8
FILING DATE: 02-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9206318.9
FILING DATE: 24-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9206372.6
FILING DATE: 23-SEP-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525004.9
FILING DATE: 07-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9610824.6
FILING DATE: 23-MAY-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB92/02240
FILING DATE: 02-DEC-1992
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/244,557
FILING DATE: 01-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: David W. Clough
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 28111/33308
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 333 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-244-557

Query Match	11.7%	Score	28.6	DB	2	Length	333
Best Local Similarity	51.1%	Pred. No.	0.79				
Matches	67	Conservative	0	Mismatches	64	Indels	0
						Gaps	0

OY	171	gaagggagaacatgataaacatgacctcaagaagcttctgtcattgatgaaagacaagc	230
Dd	113	AACAGCACCACCGGCAAAAGCCGCCCAACTCATGATTATGATGTGATGAACGGCGCCTCAG	172
OY	231	tgcagcgctgagc	241
Dd	173	GGGTCCCTTGAC	183

RESULT 5
 US-08-826-611-1/C
 ; Sequence 1, Application US/08826611
 ; Patent No. 6031154
 GENERAL INFORMATION:
 APPLICANT: Bennett, Alan B.
 APPLICANT: Kanayama, Yoshiro
 TITLE OF INVENTION: Fructokinase Genes and Their Use in
 TITLE OF INVENTION: Metabolic Engineering of Fruit Sweetness
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/826,611
 FILING DATE: 05-APR-1997
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Bastian, Kevin L.
 REGISTRATION NUMBER: 34,774
 REFERENCE/DOCKET NUMBER: 023070-077400US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1907 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

```

1 MOLECULE TYPE: CDNA
2
3 FEATURE:
4
5 NAME/KEY: CDS
6 LOCATION: 271..1314
7
8 OTHER INFORMATION: /product="tomato Frk1"
9
10 US-08-826-611-1

```

Query Match	11.6%	Score 28.4	DB 3	Length 1907
Best Local Similarity	53.6%	Pred. No. 2.1		
Matches 59; Conservative	0	Mismatches	51	Indels 0; Gaps 0;

[illegible]

RESULT 6
US-07-817-917-2/c
; Sequence 2, Application US/07817917
; Patent No. 5405763

GENERAL INFORMATION: Masako NISHIZAWA
APPLICANT: Masako NISHIZAWA
APPLICANT: Hideo OKAWA
APPLICANT: Masataru KANAOKA
TITLE OF INVENTION: ESTERASE
TITLE OF INVENTION: GENE ENCODING ASYMMETRICALLY ACTIVE
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:

ADDRESS: Birch, Stewart, Kolasch & Bir
 STREET: 301 No. 5405763th Washington Str
 CITY: Falls Church
 STATE: VIRGINIA
 COUNTRY: U.S.A.
 ZIP: 22046-3487
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 5.25 inch, 500 Kb
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Wordperfect
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/817.917
 FILING DATE: 19920109
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 001537/1991
 FILING DATE: 10-JAN-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Svensson, Leonard R.
 REGISTRATION NUMBER: 30,330
 REFERENCE/DOCKET NUMBER: 20-2941P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-241-1300
 TELEFAX: 703-241-2848
 TELEX: 248345
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2174 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: genomic DNA
 ORIGINAL SOURCE:
 ORGANISM: *Arthrobacter globiformis*
 STRAIN: SC-6-98-28 (Ferm BP-3618)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: from 1 to 1125
 IDENTIFICATION METHOD: E

US-07-817-917-2

Query Match	11.58;	Score 28.2;	DB 1;	Length 2174;
Best Local Similarity	64.68;	Pred. No. 2.7;		
Matches 42;	Conservative 0;	Mismatches 23;	Indels 0;	Gaps 0

Qy	gctctg	239
Db	2031	GCCTG 2027

```

RESULT 7
US-08-207-904-1
Sequence 1, Application US/08207904
Patent No. 5477002
GENERAL INFORMATION:
APPLICANT: Tuttle, AnnMarie
APPLICANT: Crossland, Lytle D.
TITLE OF INVENTION: Anthr- Specific cDNA Sequences, Genomic
TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
City: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/207,904
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/908,242
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lazear, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: CGC 1624
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8615
TELEFAX: (919)541-8689
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1542 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Nicotiana tabacum
INDIVIDUAL ISOLATE: Ant32
FEATURE:
NAME/KEY: CDS
LOCATION: 66..1412
US-08-207-904-1
Query Match 11.3%; Score 27.8; DB 1; Length 1542;
Best Local Similarity 59.5%; Pred. No. 3.2;

```

[illegible]

```

RESULT      8
US-08-207-904-16
; Sequence 16, Application US/08207904
; Patent No. 5477002
; GENERAL INFORMATION:
; APPLICANT: Tuttle, AnnMarie
; APPLICANT: Crossland, Lyle D.
; TITLE OF INVENTION: Anthr-Specific cDNA Sequences, Genomic
; TITLE OF INVENTION: DNA Sequences and Recombinant DNA Sequences
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/207,904
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/908,242
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: CGC 1624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8615
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3706 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Nicotiana tabacum
; INDIVIDUAL ISOLATE: Ant32 genomic clone
; IMMEDIATE SOURCE:
; CLONE: PCIB950
; FEATURE:
; NAME/KEY: TATA_signal
; LOCATION: 1971..1975
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2076..3422
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2009
; OTHER INFORMATION: /note="Putative transcription
; OTHER INFORMATION: start site"
US-08-207-904-16

```



```

11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1
```

RESULT 12
 US-09-091-590A-2/c
 Sequence 2, Application US/09091590A
 Patent No. 6242574
 GENERAL INFORMATION:
 APPLICANT: Nielsen, Klaus
 APPLICANT: Kroil Kristensen, Anne
 APPLICANT: Brunstedt, Janne
 TITLE OF INVENTION: Anti-Microbial proteins
 FILE REFERENCE: S-137-1101/MA/A/SSS/PCT
 CURRENT APPLICATION NUMBER: US/09/091,550A
 CURRENT FILING DATE: 1999-05-06
 PRIOR APPLICATION NUMBER: PCT/EP96/05765
 PRIOR FILING DATE: 1996-12-20
 PRIOR APPLICATION NUMBER: GB 9526238.2
 PRIOR FILING DATE: 1995-12-21
 NUMBER OF SEQ ID NOS: 35
 SOFTWARE: Patentin version 3.0
 SEQ ID NO: 2
 LENGTH: 587
 TYPE: DNA
 ORGANISM: Beta vulgaris
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (129)..(374)
 US-09-091-590A-2

Query Match	11.0%;	Score 27;	DB 4;	Length 587;
Best Local Similarity	46.9%;	Pred. No. 3.8;		
Matches 84;	Conservative	0;	Mismatches 95;	Indels 0;
			Gaps	0;

Oy	26	atccctcgcgactaacctctcgaatcagaagcctctcctctaaagtcacaggaacatgcc	85
Db	260	AACGCGTTCTATCAACCGCTTTTAAATACTCGCCGCGGTCTTGTTCGTGATGCTAT	201
Oy	86	ttgcactctgagcactgcgccttctacagagatttgcctctgagaaagccctgagatct	145
Db	200	TATCTGCTTGTAAAGCACCAGATCTTCTCAAGACTTGTCCCTTTAAAGTCTTCATAT	144
Oy	146	gctaatactatcacatctaggtgctcgtgaagggaaacagatgaagaacatgacctcaaga	204
Db	140	AGTTGGTTCATTTCTTTCGCAATGTTTGAAGCGGAATATTTGAGCAACGAAGAAGACTAAGA	82

RESULT 13
 US-09-276-531-52
 Sequence 52. Application US/09276531
 Patent No. 618368
 GENERAL INFORMATION:
 APPLICANT: Bandman, Olga
 APPLICANT: Lal, Preeti
 APPLICANT: Hillman, Jennifer L.
 APPLICANT: Yue, Henry
 APPLICANT: Reddy, Roopa
 APPLICANT: Baugher, Karl J.
 APPLICANT: Baughn, Mariah R.
 TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
 RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
 NUMBER OF SEQUENCES: 134
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 STREET: 3174 PORTER DRIVE
 CITY: PALO ALTO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2


```

TELEPHONE: 703-521-2297
TELEFAX: 703-685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1441 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-136-277-18

Query Match          11.0%; Score 27; DB 1; Length 1441;
Best Local Similarity 53.3%; Pred. No. 5.9;
Matches 57; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY      34 gcatcactctctgtgagtcagagccctctctcctaagtacgagggaactgccttgcctaact 93
        ||||| ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      319 GGATGTGGTCTGACCCCTGGGCCCTGGTGTTAGGCACAATGGAAACCTTGCGATGGCCCC 260

OY      94 tttgaccctggcccttacctacagcagatttttgtcttgsgaaagccctggg 140
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      259 TGAAGCTGTCCGCTTTAAGGCCCCCAGGGGGTGTCTGTGGGCCCTTCGTGTG 213
```

RESULT 15
 US-08-479-403-18/c
 : Sequence 18, Application US/08479403
 : Patent No. 5869039
 :
 : GENERAL INFORMATION:
 : APPLICANT: MANDEL, Jean-Louis
 : APPLICANT: AUBOURG, Patrick
 : APPLICANT: MOSSER, Jean
 : APPLICANT: SARDE, Claude
 : TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND
 : TITLE OF INVENTION: CORRESPONDING PROTEIN
 : NUMBER OF SEQUENCES: 23
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Young & Thompson
 : STREET: 745 South 23rd Street
 : CITY: Arlington
 : STATE: VA
 : COUNTRY: USA
 : ZIP: 22202
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08-479,403
 : FILING DATE: 07-JUN-1995
 : CLASSIFICATION: 514
 :
 : ATTORNEY/AGENT INFORMATION:
 : NAME: PATCH, Andrew J
 : REGISTRATION NUMBER: 32,925
 : REFERENCE/DOCKET NUMBER: B2272DIV
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 703-521-2297
 : TELEFAX: 703-685-0573
 : TELEEX: 248425 EMBON
 :
 : INFORMATION FOR SEQ ID NO: 18:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1441 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 :
 : MOLECULE TYPE: DNA (genomic)
 :
 : US-08-479-403-18

Best Local Similarity	53.3%;	Pred. No. 5.9;			
Matches	57; Conservative	0; Mismatches	50; Indels	0; Gaps	0;

34 gcatlactctctgagtcagagcctcttctctctlaagtcacgaggaactgccttgtaact 93

Db 319 GGATTGTTGCTCTGACCCCTGGGCGCTGTTGTTCTAGGCACTGGAACCTTGGCGATGGCCCC 260

94 tgtgacctgccttactcagcagtttltgttctggaagccctgg 140

Db 259 TGAGACTGTGGCTTTAGGCCCCCGGGGTGGTGTCTGGGGCCCCCTGGTG 213

Search completed: September 26, 2002, 04:22:19
Job time: 8869 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 01:51:15 : Search time 3532.1 Seconds
(without alignments)
936.201 Million cell updates/sec

Title: US-09-874-390-1_COPY_1_245

Perfect score: 245

Sequence: 1 attactcggttctctatt.....caagctgacgcctgcgaag 245

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	182	74.3	904	10	BG469649 602534110
2	179	73.1	671	10	BG750604 602708441
3	175	71.4	301	9	AA295520 EST100689
4	174	71.0	826	10	BG821370 BG821370
5	172	70.2	912	10	BG469141 BG469141
6	170	69.4	433	9	AA315469 AA315469
7	170	69.4	871	10	BG386293 BG386293
8	158	64.5	955	10	BG328280 BG328280
9	157.8	64.4	503	9	AM603640 AM603640
10	156.8	64.0	736	10	BE617113 BE617113
11	151.8	62.0	300	9	AU099084 AU099084
12	141	57.6	297	9	AA345780 AA345780
13	138.2	56.4	850	10	BG82181 BG82181
14	134	54.7	717	9	AM603669 AM603669
15	131	53.5	805	10	BG469804 BG469804
16	127	51.8	702	10	BG248304 BG248304
17	124.4	50.8	809	10	BE872889 BE872889

18	122	49.8	583	9	AM964294
19	106	43.3	978	10	BG171515
20	105.4	43.0	882	10	BG171085
21	82.8	33.8	763	10	BT826675
22	80	32.7	1234	10	BG824793
23	73.4	30.0	830	10	BE868512
24	72	29.4	920	10	BG386247
25	72	29.4	1003	10	BG385575
26	61	24.9	864	10	BG821420
27	61	24.9	872	10	BG747491
28	56.8	23.2	629	10	BG167918
29	35.6	14.5	954	10	BE671130
30	35	14.3	451	12	A2480015
31	34.8	14.2	857	10	BG721017
32	34.6	14.1	778	12	A2093362
33	34.6	14.1	881	12	BH163660
34	34.4	14.0	403	9	A1484837
35	34.4	14.0	567	9	A1486191
36	34.4	14.0	577	9	AM648029
37	34	13.9	1333	10	BM476682
38	33.8	13.8	1813	10	BE964358
39	33.4	13.6	456	12	AQ378568
40	33.4	13.6	495	12	A2508588
41	33.4	13.6	734	10	BG066626
42	33.4	13.6	885	12	A2201927
43	33.2	13.6	446	9	AA703887
44	33.2	13.6	536	9	BM328625
45	33.2	13.6	543	12	AQ303351

ALIGNMENTS

RESULT 1
LOCUS BG469649 904 bp mRNA linear EST 21-MAR-2001
DEFINITION 602534110F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4661691 5',
LOCUS BG469649
ACCESSION BG469649.1 GI:13401924
VERSION BG469649
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: LITCM1461 row: c column: 04
High quality sequence stop: 604.
Location/Qualifiers
1..904
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4661691"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pGBM7; Site:1: XhoI; Site:2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAC(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in

REFERENCE
AUTHORS NIH-MGC http://imgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov

the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)".

Query Match 73.1%; Score 179; DB 10; Length 671;
Best Local Similarity 99.5%; Pred. No. 2.5e-44;
Matches 190; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Query Match 74.3%; Score 182; DB 10; Length 904;
Best Local Similarity 100.0%; Pred. No. 3.3e-45;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 ctcaagtcagggaatgcctctgtactgtgactgccccttactcaagattttg 123
|||||
Db 2 CTCGAATCAGCGGAGACTGCCCTGTCTACTGTGACTGCTTACTCAGCACTTTTG 61
QY 124 ttctgggaagccctgggattctgctaatacctacactgtagtgctgaaggaacaga 183
|||||
Db 62 TTCGGGAAGCCCTGGGATTCTGCTAATACCTACTAGTGCGTGAAGGAAACAGA 121
QY 184 tgaagaacatgacctcaagaagcttctgtcaatagagaagcaagctgacgtgcaa 243
|||||
Db 122 TGAGAACATGACCTCAAGGAGCTTCTGTCATGAGAGAACCAAGCTGACGCTGGCAA 181
QY 244 ag 245
||
Db 182 AG 183

RESULT 2

LOCUS BG750604 671 bp mRNA linear EST 15-MAY-2001
DEFINITION 602708441F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4844843 5',
mRNA sequence.
ACCESSION BG750604
VERSION BG750604.1 GI:14061257
KEYWORDS EST.

SOURCE

human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 671)

TITLE NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LICM1683 row: b column: 12
High quality sequence stop: 671.
Location/Qualifiers

FEATURES

source

1. 671
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4844843"
/clone_1b="NIH_MGC_43"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pORF7; Site: 1; XhoI; Site: 2;
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. |"

BASE COUNT

165 a 161 c 166 g 179 t

QY 55 gccctctctcctaagtcagggaatgcctctgtactgtgactgccccttactcag 114
|||||
Db 2 GCCCTCTCTCTCTAAGACAGGAGAACTGCTTCTGCTGTGACCTGCTTACTCAG 61
QY 115 cagttttgtcttggaagccctgggattctgctaatacctacactgtagtgctgaag 174
|||||
Db 62 CAG-TTTTGTCTTGGAAGCCCTGGGATTCTGCTAATACCTACTAGTGCTCTAAG 120
QY 175 ggaacagatgaagaatgacctcaagaagcttctgtcaatagagaagcaagctgac 234
|||||
Db 121 GGAACAGATGAAGACATGACCTCAAGGAGCTTCTGTCATGAGAGAACCAAGCTGAC 180
QY 235 gccctggcaag 245
|||||
Db 181 GCCTGGCAAG 191

RESULT 3

LOCUS AA295520 301 bp mRNA linear EST 18-APR-1997
DEFINITION EST100689 Pancreas tumor I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA295520
VERSION AA295520.1 GI:1947854
KEYWORDS EST.

SOURCE

human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 301)

AUTHORS

Adams M.D., Kerlavage A.R., Fleischmann R.D., Feldner R.A., Bult
, C.J., Lee N.H., Kirkness E.F., Weinstock K.G., Gocayne J.D., White
, O., Sutton G., Blake J.A., Brandon R.C., Man-Wai C., Clayton R.A.,
Cline T.R., Cotton M.D., Earle-Hughes J., Fine L.D., Fitzgerald
, L.M., Fitzhugh W.M., Fitchman J.L., Geoghagen N.S., Glodok A.,
Gnehm C.L., Hanna M.C., Hedblom E., Hinkle P.S., Jr., Kelley J.M.,
Kelley J.C., Liu L.-I., Marmaros S.M., Merrick J.M.,
Moreno-Palauques R.F., McDonald L.A., Nguyen D.T., Pelligrino S.M.,
Phillips C.A., Ryder S.E., Scott J.L., Sauder D.M., Shirley R.,
Small K.V., Spriggs T.A., Uteyback T.R., Weidman J.F., Li Y.,
Bednarik D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J.,
Dimke D., Feng D.-F., Ferris A., Fischer C., Hastings G.A., He W.W.,
Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K., Kozak D.L.,
Kunsch C., Hungjun J., Li H., Weisener P.S., Olsen H., Raymond L.,
Wei Y.F., Wang J., Xu C., Yu G.L., Ruben S.M., Dillon P.J., Fannon
, M.R., Rosen C.A., Haseltine W.A., Fields C., Fraser C.M. and
Venter J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)

JOURNAL

MDLINE
COMMENT Other-ESTs: THC168761
Contact: Kerlavage, AR
Bioinformatics

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/ngi/ngi.html)
Seq primer: M13 Reverse.

FEATURES

source

1. 301
/organism="Homo sapiens"
/db_xref="ATCC (inhost):190965"
/db_xref="taxon:9606"

BASE COUNT	ORIGIN
82 a	66 c 75 g 77 t 1 others

Query Match	71.4%;	Score 175;	DB 9;	Length 301;
Best Local Similarity	99.4%;	Pred. No. 3.4e-43;		
Matches 175; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

```
QY 70 gtcacggagactgcccctgtcgaacttgyaacccyccccctaactcaagacaglttttgctcg 129
    |||||||
Db 1 gTCAcGGGAActGCCCTTGACTTGAGACCCGCTTAActCACCAGTTTGTGTCGG 60
```

Qy 130 gaagccctggtatttcgtctaatacctatcacigttagtgcgtgaaggaaacaga tgaaga 189
|||||
|||
Db 61 GAAGCCCTGGGATTTCGTCTAATACCTATCACCTGTAGTGCTCGAAGGAAACANATGAAGA 120

```

QY 190 acatgaccctcaaggagcttcctgtccatgagagaccacgaagtgcgcctggcaaa 245
    |||||
    |||||
    |||||
Db 121 ACATGACCTCAAGGAGCTTCCTGTCAATGAGAGAACCACGCTGACGCTGGCAAA 176

```

RESULT	4				
BG821370					
LOCUS					
BC821370					
826 bp					
EMBL					
14002					
ECM					
22-MAY-2001					

DEFINITION	B02748B14F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:8664634 5'
ACCESSION	BG821370
VERSION	GI:14169957
DESCRIPTION	mRNA sequence.

NAMEWORDS	SOURCE	ORGANISM
ESI.	human.	Homo sapiens
		Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:

REFERENCE
1 (bases 1 to 826)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found at <http://www.ncbi.nlm.nih.gov/Genbank/Genbank.html>

http://image.lnl.gov
Plate: LLCM1728 row: k column:
High quality sequence stop: 823.

```
source
1. .826
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```

/clonelib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"

```

ECORI; cDNA made by oligo-dT priming. Directionally cloned into EcorI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1 kb library constructed by ligation.

BASE COUNT 217 a 197 c 205 d 207 t
line laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)

Query Match	71.08;	Score 174;	DB 10;	Length 826;
-------------	--------	------------	--------	-------------

Best Local Similarity 100.0%; Pred. NO. 9.2e-43;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1/2 cagcggaaacgcccctgctactgtgacacgcccctactcagcagtttctgtcggga 13
|||||
Dy 2 CAGGGGAACGCCCCCTTGCTACTTGTGACCTGCCCTTACTCAGCAGTTTCTGTCTGGGA 61

132 agccctgggattctgctaataccatcaactgtaggtgcctgaagggaacacagatgaagaac 134
 |||||
 62 AGCCCTGGGATTCTGCTAATACCTATCATCTGTAGGTGCTGAAGGGAACAGATGAAGAAC 122

Dy 192 atgaccccaaggagcctcccgccaacgaagaaccacgcgcgcgcggccaad 24
| | | | |
Db 122 atgacctcaaggagcttcctgtcatgagaaaccaagtgcgcgcggccaag 175

RESULT	5			
LOCUS	BG469141	912 bp	mRNA	linear
	EST	21-MAR-2001		

Accession	Version	Accession	Version
MG469141	MG469141.1	GI:13401416	

LOC	SOURCE	ORGANISM
human.		Homo sapiens
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE

1 (bases 1 to 912)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/INR at:

http://image.liml.gov
 Plate: LLCM1418 row: b column: 08
 High quality sequence stop: 705.
 Location/Qualifiers

```

SOURCE
1. 512
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cclone="IMAGE:4645159"

```

```

/culture_id= NH100013
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="organ: colon; Vector: pOTB7; Site 1: XhoI; Site 2:

```

library was made by large and primary. directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in

California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).¹

Query Match 70.2%; Score 172; DB 10; Length 912;

Matches	172;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
74	cgggaactgcccctgtactgtgtgactgtgccccttactcagcagattttgtctctggaag	133							

Dd 2 CGGGAACACGCCCTTGCTACTTGTGACCTGCCCCCTTACTCAGCAGTTTTTGTCTGGGAG 61

Qy 134 ccctggaattctgctaatacctatacctglaagtgctgaaggaaacagatgaagaacat 193

Db 62 CCCTGGATCTGCTAATACCTATGCTGATGCTGTAAGGGAAGAAACATGAAACAT 121

QY 194 gaactcaaggagcttctctgcaatgagaagcaagctgacgcttggcaag 245
|||||

Db 122 GACCTCAAGAGAGCTTCTGCTCAATGAGAGAGCAAGCTGACGCTGGCAAG 173

RESULT 6
AA315469
LOCUS
DEFINITION
EST187245 Colon carcinoma (HCC) cell line II Homo sapiens CDNA 5'

ACCESSION
AA315469
VERSION
AA315469.1 GI:1967798
KEYWORDS
EST.
SOURCE
Homo sapiens
human.

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wei,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,
L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodok,A.,
Guelm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
Kelly,D.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Weisner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)

JOURNAL
MEDLINE
96026280

COMMENT
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers
1. 433
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):110433"
/db_xref="taxon:9606"
/clone_lib="Colon carcinoma (HCC) cell line II"
/tissue_type="colon"
/cell_type="KM12C"
/cell_line="KM12C(HCC)-parental human colon carcinoma
;Dukes B2
/note="Organ: colon; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 112 a 101 c 103 g 112 t 5 others
ORIGIN

Query Match 69.4%; Score 170; DB 9; Length 433;
Best Local Similarity 99.4%; Pred. No. 1.3e-41;
Matches 170; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 75 gggaagcctcctgactctgtgacgtcccttactcagcagtttctgttgggaagc 134

Db 1 GGGAGCTGCCCTTGCTACTGTGTNACCTGCCCTTACTACAGATTTTGTGGAGC 60
|||||

QY 135 ccttgagattctgtcaatcacctatcacctgtgctgtaaggaacagatgaacatg 194
|||||

Db 61 CTTGGGATTTCTGCTAATACCTATGCTAGTGTCTCTAGAGGAACAGATGAACATG 120
|||||

QY 195 acctcaaggagcttctctgcaatgagaagcaagctgacgcttggcaag 245
|||||

Db 121 ACCTCAAGAGAGCTTCTGCTCAATGAGAGAGCAAGCTGACGCTGGCAAG 171
|||||

RESULT 7
BG386293
LOCUS
DEFINITION
602455486F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583948 5',
mRNA sequence.

ACCESSION
BG386293
VERSION
BG386293.1 GI:13279739
KEYWORDS
EST.
SOURCE
human.
Homo sapiens

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 871)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LNCMI308 row: k column: 21
High quality sequence stop: 736.
Location/Qualifiers
1. 871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:4583948"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 231 a 212 c 215 g 213 t
ORIGIN

Query Match 69.4%; Score 170; DB 10; Length 871;
Best Local Similarity 99.5%; Pred. No. 1.6e-41;
Matches 181; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 64 ctcaagtcacggaagctcctgtctactgtgtacgtcccttactcagcagttttg 123
|||||

Db 2 CTCTAAGTCAAGGAGAACTGCCCTTGCTACTGTGTACTCGCCTTACTACAGAG-TTTTG 60
|||||

QY 124 ttctggaagccttggattctgtaataacatcacctgtagtgtgtaagggaacaga 183
|||||

Db 61 TTCTGGGAAGGCCCTGGATCTCTGCTAATACCTATGCTAGTGTGTAAGGAAGAGA 120
|||||

QY 184 tgaagaacatgactcaagagcttctgttcaatgagaagcaagctgacgcttgcaa 243
|||||

Db 121 TGAAGACATGACCTCAGAGAGCTTCTCATATGAGAGACCAAGCTGAGCGCTGGCAA 180
OY 244 ag 245
11
Db 181 AG 182

RESULT 8
BG328280 955 bp mRNA linear EST 27-FEB-2001
LOCUS 602427256p1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4546813 5',
DEFINITION mRNA sequence.
ACCESSION BG328280
VERSION BG328280.1 GI:13134627
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 955)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM1233 row: P column: 14
High quality sequence stop: 718.
Location/Qualifiers
1..955
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4546813"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 245 a 230 c 246 g 233 t 1 others
ORIGIN

Query Match 64.5%; Score 158; DB 10; Length 955;
Best Local Similarity 98.9%; Pred. No. 8e-38;
Matches 180; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

OY 64 ccttaagtaagggaactccttgctacttgctgaccccttacttaagcaagttttg 123
11
Db 2 ccttaagtaagggaactccttgctacttgctgaccccttacttaagcaagttttg 59
11
OY 124 ttctgggaagccctggatctctgtaataactacactgtagtgctggaaggaacaga 183
11
Db 60 ttctgggaagccctggatctctgtaataactacactgtagtgctggaaggaacaga 119
11
OY 184 tgaagaacatgacctaagaagacttctcgtcaatgagaagaccagctaacgcttgcaa 243
11
Db 120 tgaagaacatgacctaagaagacttctcgtcaatgagaagaccagctaacgcttgcaa 179
11
OY 244 ag 245
11
Db 180 AG 181

RESULT 9
AM603640/c 503 bp mRNA linear EST 23-MAR-2000
LOCUS CM0-CN0039-260100-166-e08 CN0039 Homo sapiens CDNA, mRNA sequence.
DEFINITION AM603640
ACCESSION AM603640
VERSION AM603640.1 GI:7308381
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 503)
AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM0&t2=CM0-CN0039-260100-166-e08&t3=2000-01-26&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 25
High quality sequence stop: 503.
Location/Qualifiers
1..503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CN0039"
/dev_stage="Adult"
/note="Organ: colon_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 145 a 109 c 112 g 137 t
ORIGIN

Query Match 64.4%; Score 157.8; DB 9; Length 503;
Best Local Similarity 98.8%; Pred. No. 7.7e-38;
Matches 159; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 85 ctgtcacttgtagactgccttacttactgagcaagtttcttggaagccctggatgc 144
11
Db 503 cttgctacttgtagactgccttacttactgagcaagtttcttggaagccctggatgc 444
11
OY 145 tcttaactcttactcactgtagtgctggaaggaacagaatgagaacatgacctcaagga 204
11
Db 443 tcttaactcttactcactgtagtgctggaaggaacagaatgagaacatgacctcaagga 384
11
OY 205 gcttcctgcaatgagaacacagctgagcgtggaag 245
11
Db 383 gcttcctgcaatgagaacacagctgagcgtggaag 343
11

RESULT 10
BE617113 736 bp mRNA linear EST 20-OCT-2000
LOCUS BE617113
DEFINITION 601441649p1 NIH_MGC_65 Homo sapiens CDNA clone IMAGE:3845677 5',
mRNA sequence.
ACCESSION BE617113
VERSION BE617113.1 GI:9888051

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Eukaryota: Eutheria: Primates: Catarrhini: Hominiidae: Homo. 1 (bases 1 to 736)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ARCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LIA9557 row: b column: 14
 High quality sequence start: 9
 High quality sequence stop: 670.
 Location/Qualifiers
 1. 736
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3845677"
 /clone_lib="NIH-MGC-65"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site: 1; NotI; Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."
BASE COUNT 191 a 170 c 183 g 191 t 1 others
ORIGIN

Query Match 64.0%; Score 156.8; DB 10; Length 736;
 Best Local Similarity 98.8%; Pred. No. 1.7e-37;
 Matches 156; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 86 ttgttacttgagctgaccttcttactcagcagtttctgtctggaagcccttgatctt 145
 |||||||
 Db 1 TTGCTACTGTGACCTGCCCTTTACTCAGCAGTTTGTTGTGGAGCCCTGGGATTCT 60

QY 146 gctaatacctatcactgtaggtctggaaggaaacagatgaagacatgacctcaaggag 205
 |||||||
 Db 61 GCTATACCTACTGCTGAGGTGCTGGAAGGAACAGATGAAGAACAATGACCTCAAGGAG 120

QY 206 ctctctgctcaatgagaagaccagctgacgcctgcaag 245
 |||||||
 Db 121 CTTCTCTGTCATGAGAACCAACGACCTGACCTGGCAAG 160

RESULT 11
AU099084 300 bp mRNA linear EST 05-APR-2001
LOCUS AU099084 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
DEFINITION HEP17444 similar to Homo sapiens
 beta-16-N-acetylglucosaminyltransferase mRNA, mRNA sequence.
ACCESSION AU099084
VERSION AU099084.1 GI:13550213
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
REFERENCE 1 (bases 1 to 300)
 Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, J., Sese, J., Hata
 H., Ota, T., Isogawa, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo
 K., Suyama, A. and Sugano, S.
 In silico mapping of the 5'-ends of human mRNAs using full-length
 enriched and 5'-end enriched cDNA libraries constructed by

JOURNAL Oligo-capping method
COMMENT Unpublished (2001)
 Contact: Yutaka Suzuki
 Institute of Medical Science, University of Tokyo
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
 Email: ysuzuki@ims.u-tokyo.ac.jp
 Suzuki, Y., Yoshimoto-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano
 S. Construction and characterization of a full length-enriched and
 a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
FEATURES
 Location/Qualifiers
 1. 300
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="HEP17444"
 /clone_lib="Sugano Homo sapiens cDNA library"
BASE COUNT 86 a 64 c 72 g 78 t
ORIGIN

Query Match 62.0%; Score 151.8; DB 9; Length 300;
 Best Local Similarity 98.7%; Pred. No. 4.6e-36;
 Matches 153; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 91 acttgtaacctgaccttaactcagcagtttctgtctggaagccctggatctgctaa 150
 |||||||
 Db 1 ACTGTGACCTGCGCCCTTACTCATTTCTTTCTGCGAAGCCCTGGGATTCTGCTAA 60

QY 151 tactatcactgtagtgcctggaaggaaacagatgaagacatgacctcaaggatctgc 210
 |||||||
 Db 61 TACTTATCTAGTGTGCTGGAAGGAACAGATGAAGAACAATGACCTCAAGAGCTTCC 120

QY 211 tgtcaatgagaagaccagctgacgcctgcaag 245
 |||||||
 Db 121 TGTCATGAGAACCAACGACCTGACCTGGCAAG 155

RESULT 12
AA345780 297 bp mRNA linear EST 21-APR-1997
LOCUS AA345780 gall bladder I Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION AA345780
ACCESSION AA345780
VERSION AA345780.1 GI:1998058
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
REFERENCE 1 (bases 1 to 297)
 Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Feldner, R.A., Bult
 C.J., Lee, N.H., Kirkness, E.F., Weissbrock, K.G., Gocayne, J.D., White
 O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Val, C., Clayton, R.A.,
 Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald
 L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A.,
 Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
 Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
 Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
 Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Sherry, R.,
 Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y.,
 Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
 Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,
 Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
 Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L.,
 Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon
 M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
 Venter, J.C.
 Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 Nature 377 (6547 Suppl.), 3-174 (1995)
JOURNAL MEDLINE
COMMENT Other ESTs: THC168761
 Contact: Kerlavage, AR
 Bioinformatics

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423

Email: arkerlavet@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES

location/Qualifiers

1..297
/organism="Homo sapiens"
/db_xref="ATCC (inhost):147350"
/db_xref="taxon:9606"
/clone_lib="Gall bladder I"
/sex="female"
/dev_stage="adult, 25 yrs"
/note="Organ: gall bladder; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 84 a 59 c 76 g 75 t 3 others
ORIGIN

Query Match 57.6%; Score 141; DB 9; Length 297;
Best Local Similarity 100.0%; Pred. No. 9.6e-33;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 105 cttctcgaagagtttctgtctggaagccctggatctgtctatacctcctgta 164
|||||
Db 1 ctttactagcagcttttctgtctggaagccctggatctgtctatacctcctgta 60

OY 165 ggtgctgaaggaacacagatgaagacatgaagcctcaagagcttctctcaatgaaga 224
|||||
Db 61 ggtgctgaaggaacacagatgaagacatgaagcctcaagagcttctctcaatgaaga 120

OY 225 ccaagctgaagccttggcaag 245
|||||
Db 121 ccaagctgaagccttggcaag 141

RESULT 13
Bg822181 850 bp mRNA linear EST 22-MAY-2001
LOCUS 602726335F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4865672.5',
DEFINITION mRNA sequence.
ACCESSION Bg822181
VERSION Bg822181.1 GI:14169768
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 850)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC

CNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LCM171 row: f column: 09
High quality sequence stop: 696.

FEATURES

location/Qualifiers

1..850
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4865672.5"
/clone_lib="NIH_MGC_15"

/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 230 a 191 c 222 g 207 t
ORIGIN

Query Match 56.4%; Score 138.2; DB 10; Length 850;
Best Local Similarity 94.7%; Pred. No. 9.4e-32;
Matches 143; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 95 gtgacctgcccttactactagcagtttctgtggaagccctggatctgtctataacc 154
|||||
Db 2 GTGACCTGCCCTTTAGTCTCCAGATAGGGTTCTGGGAAGCCCTGGATCTCTCATATACC 61

OY 155 tatcactgtaggtgctggaaggaacagatgaagacatgaagcctcaagagcttctgtc 214
|||||
Db 62 TATCACTGTAGGTGCTGGAAGGAACAGATGAAGACATGACCTCAAGAGCTTCTCTGTC 121

OY 215 aatgagaagaccgaagctgaagccttggcaag 245
|||||
Db 122 AATGAGAAGACCAGACTGACGCTTGCAAG 152

RESULT 14
AM603669 717 bp mRNA linear EST 23-MAR-2000
LOCUS CM0-CN0039-280100-168-e02 CN0039 Homo sapiens cDNA, mRNA sequence.
DEFINITION
ACCESSION AM603669
VERSION AM603669.1 GI:7308410
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 717)
AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM0<2=CM0-CN0039-280100-168-e02<3=2000-01-28<4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 630.

FEATURES

location/Qualifiers

1..717
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CN0039"
/dev_stage="Adult"
/note="Organ: colon_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 194 a 141 c 168 g 214 t
ORIGIN

Query Match 54.7%; Score 134; DB 9; Length 717;
Best Local Similarity 99.3%; Pred. No. 1.8e-30;
Matches 145; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 100 ctgaccttactcaagcagtttctgtctgggaagccctgggattctgctaatacctatca 159
DB 14 CTGCCCTTACTCAGCAG-TTTTGTCTGGGAAGCCCTGGGATTCTCTAATACCTATCA 72
QY 160 ctgtaggtctgaaggaacacagatgaacatgacctcaaggagctctctgtaatga 219
DB 73 CTGTAGTGTCTGAAGGAACAGATGAAGACATGACCTCAAGGAGCTTCTGCAATGA 132
QY 220 gaagaccaagctgacgcttgcaag 245
DB 133 GAAGACCAAGCTGACGCTGGCAAG 158

RESULT 15

BG469804

LOCUS 805 bp mRNA linear EST 21-MAR-2001

DEFINITION 602534325F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4661670 5',

mRNA sequence.

ACCESSION BG469804

VERSION BG469804.1 GI:13402079

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 805)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabps-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: NIH Intramural Sequencing Center

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: L1CM1461 row: b column: 07

High quality sequence stop: 623.

FEATURES

source

1..805
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4661670"
/clone.lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the Laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT

205 a 205 c 191 g 204 t

ORIGIN

Query Match 53.5%; Score 131; DB 10; Length 805;

Best Local Similarity 94.4%; Pred. No. 1.5e-29;

Matches 169; Conservative 0; Mismatches 5; Indels 5; Gaps 3;

QY 61 tctcttaagtcacgggaactgcctgtacttg-tgacctgaccttactcagcagtt 119

DB 3 TCTCTTAAGTCAAGGGAACCTGCTTACTGTGACCTGCCCTTACTACAGCAGTT 62
QY 120 ttgtctctgggaagccctgggattctgctaatacctatcaactgtagtctgaaggaa 179
DB 63 TATGTTCTGGGAAGCCCTGGGATTCTCTAATACCTATGAGTGTGAAGGAA 122
QY 180 ca--gatgaagaacatgacct--caaggagcttctgtaatgagaagaccagctgac 234
DB 123 CAGGATGAAGAACATGACCTTCAAGGAGGCTCTCTGTAATGAGAAGACCAAGCTGAC 181

Search completed: September 26, 2002, 02:54:54
Job time: 3819 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 02:54:54 ; Search time 3532.1 Seconds

(without alignments)
726.033 Million cell updates/sec

Title: US-09-874-390-1_COPY_246_435

Perfect score: 190

Sequence: 1 acattaaagagagagcctgaa.....caggagagaagctactaaag 190

Scoring table: IDENTITY_NUC

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190	100.0	583	9	AM964294 EST376367
2	190	100.0	702	10	BG248304 602400340
3	190	100.0	826	10	BG821370 602724814
4	190	100.0	864	10	BG821420 602724876
5	190	100.0	871	10	BG386293 602455486
6	188.4	99.2	850	10	BG822181 602726335
7	188.4	99.2	912	10	BG469141 602511125
8	188.4	99.2	955	10	BG328280 602427256
9	188.4	99.2	1003	10	BG385575 602453872
10	188	98.9	433	9	AA315469 EST187245
11	187.4	98.6	830	10	BE868512 601444525
12	186.8	98.3	912	10	BE871910 601447936
13	178	93.7	736	10	BE617113 601441649
14	178	93.7	763	10	B1826675 603077437
15	178	93.7	872	10	BG747491 602704606
16	178	93.7	920	10	BG386247 602455439
17	178	93.7	978	10	BG171515 602322061

18	176.4	92.8	629	10	BG167918
19	176.4	92.8	671	10	BG750604
20	176.4	92.8	882	10	BG171085 602324130
c 21	169.4	89.2	337	9	AM360893 PM1-CT024
22	161.4	84.9	904	10	BG466649 602534110
23	149.8	78.8	297	9	AA345780 EST51885
24	145	76.3	300	9	AU099084 A0099084
25	111.6	58.7	805	10	BG469804 602534325
26	111.4	58.6	301	9	AA295520 EST100689
27	102.2	53.8	766	10	B1765410 603050445
28	86.6	45.6	431	10	B1337337 AR089D07P
29	86.6	45.6	439	10	B284104 SS284104 PO
30	84.8	44.6	582	10	B138942 363673 MA
31	71.6	37.7	259	10	BE748971 MR0-BN011
32	47	24.7	809	10	BE872889 601450737
33	45	23.7	875	10	BG470202 602537311
34	40	21.1	583	10	BF041857 BP250024B
c 35	35.2	18.5	424	10	AO196782 CIT-HSP-2
36	34.2	18.0	735	10	B1934209 EST554098
37	33.4	17.6	1100	12	CNS00BL6 AL057316 Drosophila
c 38	33	17.4	854	12	AG116458 Pan trogl
39	32.6	17.2	651	12	AG045807 Pan trogl
40	32.2	16.9	561	10	B1199475 B1199475
c 41	32	16.8	611	12	AZ364585 IM011019
c 42	31.6	16.6	299	10	BF545498 UT-R-Cl-3
c 43	31.6	16.6	392	10	BG998048 PM0-HT091
44	31.4	16.5	681	12	BH261745 CH230-420
c 45	31.4	16.5	914	12	CNS06WYO AL419062 T7 end of

ALIGNMENTS

RESULT 1
LOCUS AM964294 583 bp mRNA linear EST 01-JUN-2000
DEFINITION EST376367 MAGE resequences, MAGH Homo sapiens CDNA, mRNA sequence.
ACCESSION AM964294
VERSION AM964294.1 GI:8154130
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 583)
Hegde, P., Ol, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt

, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
Quackenbush, J.
Assessment of gene expression patterns in a model of colon tumor

metastasis using a 19,200 element cDNA microarray
Unpublished (2000)

JOURNAL Contact: John Quackenbush

COMMENT The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208

Email: johnq@tigr.org
Plate: 200

Seq primer: Reverse.

Location/Qualifiers

1. 583
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGH"
/note="vector: pBluescriptKm"

BASE COUNT 153 a 135 c 140 g 155 t

ORIGIN

Query Match 100.0%; Score 190; DB 9; Length 583;
Best Local Similarity 100.0%; Pred. No. 1.7e-47;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 atattaagaagagagcctgaactgtcttccttgagacatttataatgataatcagaataaccc 60
    |||||||
DB 123 ATATTAAAGAGGAGCGCTGAAGCTGTTCTTGACATCTTATGAATGTCAGAAAATACCTT 182
QY 61 ttggaaggttaagaagatcagggagacatgtgtgttcacatttgcgcacaggaacccgcc 120
    |||||||
DB 183 TTGGAGGGTTAGAGATCAGGGGACATGTTGTTACATTGCTGCCACGCAACCCGCC 242
QY 121 agcttcacttggaacagaatcacgcctgtgtgaagaagatcaccctaaacagagagaa 180
    |||||||
DB 243 AGCTTCACCTTGGAACAGATCAGCGCTTGTAAGAGATCATCCCTAAGCAGAGAGAA 302
QY 181 gctactaaag 190
    |||||||
DB 303 GCTACTAAAG 312

RESULT 2
BG248304 702 bp mRNA linear EST 13-FEB-2001
LOCUS 602400340F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546017 5',
DEFINITION mRNA sequence.
ACCESSION BG248304
VERSION BG248304.1 GI:12758119
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 702)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
            Plate: LNCM1231 row: o column: 10
            High quality sequence stop: 702.
FEATURES
    source
        1..702
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4546017"
        /clone_lib="NIH_MGC_15"
        /tissue_type="adenoacarcinoma cell line"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: colon; Vector: pORF7; Site:1: XhoI; Site:2:
        EcoRI; CDNA made by oligo-dT priming. Directionally
        cloned into EcoRI/XhoI sites using the following 5'
        adaptor: GGCAAGG(G). Size-selected >500bp for average
        insert size 1.8kb. Library constructed by Ling Hong in
        the laboratory of Gerald M. Rubin (University of
        California, Berkeley) using Zap-cDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 186 a 163 c 175 g 178 t
ORIGIN

Query Match 100.0%; Score 190; DB 10; Length 702;
Best Local Similarity 100.0%; Pred. No. 1.8e-47;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atattaagaagagcctgaactgtcttccttgagacatttataatgataatcagaataaccc 60
    |||||||
DB 128 ATATTAAAGAGGAGCGCTGAAGCTGTTCTTGACATCTTATGAATGTCAGAAAATACCTT 187
QY 61 ttggaaggttaagaagatcagggagacatgtgtgttcacatttgcgcacaggaacccgcc 120
    |||||||

```

```

DB 188 TTGGAGGGTTAGAGATCAGGGGACATGTTGTTACATTGCTGCCACGGAACACCGCC 247
QY 121 agcttcacttggaacagaatcacgcctgtgtgaagaagatcaccctaaacagagagaa 180
    |||||||
DB 248 AGCTTCACCTTGGAACAGATCAGCGCTTGTAAGAGATCATCCCTAAGCAGAGAGAA 307
QY 181 gctactaaag 190
    |||||||
DB 308 GCTACTAAAG 317

RESULT 3
BG821370 826 bp mRNA linear EST 22-MAY-2001
LOCUS 602724814F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4864634 5',
DEFINITION mRNA sequence.
ACCESSION BG821370
VERSION BG821370.1 GI:14168957
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 826)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: LNCM1728 row: k column: 03
            High quality sequence stop: 823.
FEATURES
    source
        1..826
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4864634"
        /clone_lib="NIH_MGC_15"
        /tissue_type="adenoacarcinoma cell line"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: colon; Vector: pORF7; Site:1: XhoI; Site:2:
        EcoRI; CDNA made by oligo-dT priming. Directionally
        cloned into EcoRI/XhoI sites using the following 5'
        adaptor: GGCAAGG(G). Size-selected >500bp for average
        insert size 1.8kb. Library constructed by Ling Hong in
        the laboratory of Gerald M. Rubin (University of
        California, Berkeley) using Zap-cDNA synthesis kit
        (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 217 a 197 c 205 g 207 t
ORIGIN

Query Match 100.0%; Score 190; DB 10; Length 826;
Best Local Similarity 100.0%; Pred. No. 1.9e-47;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atattaagaagagcctgaactgtcttccttgagacatttataatgataatcagaataaccc 60
    |||||||
DB 176 ATATTAAAGAGGAGCGCTGAAGCTGTTCTTGACATCTTATGAATGTCAGAAAATACCTT 235
QY 61 ttggaaggttaagaagatcagggagacatgtgtgttcacatttgcgcacaggaacccgcc 120
    |||||||
DB 236 TTGGAGGGTTAGAGATCAGGGGACATGTTGTTACATTGCTGCCACGGAACCCGCC 295
QY 121 agcttcacttggaacagaatcacgcctgtgtgaagaagatcaccctaaacagagagaa 180
    |||||||
DB 296 AGCTTCACCTTGGAACAGATCAGCGCTTGTAAGAGATCATCCCTAAGCAGAGAGAA 355

```

QY 181 gctactaaag 190
|||||
Db 356 GCTACTAAG 365

RESULT 4
BG821420 864 bp mRNA linear EST 22-MAY-2001
LOCUS 602724876F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4864554 5',
DEFINITION mRNA sequence.
ACCESSION BG821420
VERSION BG821420.1 GI:14169007
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 864)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1728 row: g column: 19
High quality sequence stop: 849.

FEATURES
source location/Qualifiers
1..864
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4864554"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 235 a 204 c 210 g 215 t

ORIGIN

Query Match 100.0%; Score 190; DB 10; Length 864;
Best Local Similarity 100.0%; Pred. No. 2e-47;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atattaagaagagcctgaacatgttccttgagcactatgaatgcgaataatcctt 60
|||||
Db 62 ATATTAAAGGAGGAGCCTGAAGAGCTTCTTGACATCTTAATGATCTGAGAAATACCTT 121

QY 61 ttggaggttagaagatcaggagacatgttgcacattgtgcacaggaacacgcgc 120
|||||
Db 122 TTGAGAGGTTAGAGATCAGGGAGCATGTGTTTCACATTTCCTGCACAGAACACCGCC 181

QY 121 agcttcacttggaacagaaatcacgcctgtgaaagagatcacccctaaagagagaa 180
|||||
Db 182 AGCTCTTCACTTGGAACAGAAATCAGCCTTGTAAGAGATCATCCCTAAGCAGAGAGAA 241

QY 181 gctactaaag 190
|||||
Db 242 GCTACTAAG 251

RESULT 5
BG386293 871 bp mRNA linear EST 12-MAR-2001
LOCUS 602455486F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583948 5',
DEFINITION mRNA sequence.
ACCESSION BG386293
VERSION BG386293.1 GI:13279739
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 871)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1308 row: k column: 21
High quality sequence stop: 736.

FEATURES
source location/Qualifiers
1..871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583948"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 231 a 212 c 215 g 213 t

ORIGIN

Query Match 100.0%; Score 190; DB 10; Length 871;
Best Local Similarity 100.0%; Pred. No. 2e-47;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atattaagaagagcctgaacatgttccttgagcactatgaatgcgaataatcctt 60
|||||
Db 183 ATATTAAAGGAGGAGCCTGAAGAGCTTCTTGACATCTTAATGATGACAGAAATACCTT 242

QY 61 ttggaggttagaagatcaggagacatgttgcacattgtgcacaggaacacgcgc 120
|||||
Db 243 TTGAGAGGTTAGAGATCAGGGAGCATGTGTTTCACATTTCCTGCACAGAACACCGCC 302

QY 121 agcttcacttggaacagaaatcacgcctgtgaaagagatcacccctaaagagagaa 180
|||||
Db 303 AGCTCTTCACTTGGAACAGAAATCAGCCTTGTAAGAGATCATCCCTAAGCAGAGAGAA 362

QY 181 gctactaaag 190
|||||
Db 363 GCTACTAAG 372

RESULT 6
BG822181 850 bp mRNA linear EST 22-MAY-2001
LOCUS 602726315F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4865672 5',
DEFINITION

ACCESSION mRNA sequence.
VERSION BG822181
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM1731 row: f column: 09
High quality sequence stop: 696.
Location/Qualifiers
1. 850
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4865672"
/clone_lib="NIH_MGC_15"
/issue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOT7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 230 a 191 c 222 g 207 t
ORIGIN
Query Match 99.2%; Score 188.4; DB 10; Length 850;
Best Local Similarity 99.5%; Pred. No. 6.1e-47;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 atattaaaggagcctgaacactgttcttgacatcttgaatgtcagaataactctt 60
DB 153 ATATTAAAGAGACCCGTGAACCTGTCCTTGACATCTTATGAATGTCAGAAATACCTT 212
QY 61 ttgagagggttaagatcagagcagatggtgtgttcacattgtctgcacaggaacacgccc 120
DB 213 GTGAGGGGTTAGAGATCAGGGGACATGCTTGTTCACATTGCTGCCACGAGAACACGCC 272
QY 121 agtctcacttgtaaacagaatcagcctgttgaaagatcattccctaagcagagagaa 180
DB 273 AGTCTTCACTTGAAACAGATCAGCCTTGTGAAGAGATCATCCTTAAGCAGAGAGAAA 332
QY 181 gctactaaag 190
DB 333 GCTACTAAAG 342
RESULT 7
LOCUS BG469141 912 bp mRNA linear EST 21-MAR-2001
DEFINITION 602511125F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4645159 5',
mRNA sequence.
VERSION BG469141
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM1418 row: b column: 08
High quality sequence stop: 705.
Location/Qualifiers
1. 912
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4645159"
/clone_lib="NIH_MGC_15"
/issue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOT7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 240 a 214 c 231 g 227 t
ORIGIN
Query Match 99.2%; Score 188.4; DB 10; Length 912;
Best Local Similarity 99.5%; Pred. No. 6.2e-47;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 atattaaaggagcctgaacactgttcttgacatcttgaatgtcagaataactctt 60
DB 174 ATATTAAAGAGACCCGTGAACCTGTCCTTGACATCTTATGAATGTCAGAAATACCTT 233
QY 61 ttgagagggttaagatcagagcagatggtgtgttcacattgtctgcacaggaacacgccc 120
DB 234 TTGAGGGGTTAGAGATCAGGGGACATGCTTGTTCACATTGCTGCCACGAGAACACGCC 293
QY 121 agtctcacttgtaaacagaatcagcctgttgaaagatcattccctaagcagagagaa 180
DB 294 AGTCTTCACTTGAAACAGATCAGCCTTGTGAAGAGATCATCCTTAAGCAGAGAGAAA 353
QY 181 gctactaaag 190
DB 354 GCTACTAAAG 363
RESULT 8
LOCUS BG328280 955 bp mRNA linear EST 27-FEB-2001
DEFINITION 602427256F1 NIH_MGC_15 Homo sapiens CDNA clone IMAGE:4546813 5',
mRNA sequence.
VERSION BG328280
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strusberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC

CNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov

Plate: LCM1233 row: p column: 14
High quality sequence stop: 718.

FEATURES

source Location/Qualifiers

1..955
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4546813"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 245 a 230 c 246 g 233 t 1 others
ORIGIN

Query Match 99.2%; Score 188.4; DB 10; Length 955;
Best Local Similarity 99.5%; Pred. No. 6.3e-47;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atattaagagagcctgaactgttccttgacatcttatgaatgacgaataactt 60
DB 182 ATATTAAAGAGAGCGCTGAACCTGTTCTTGACATCTTATTAATGTCAGAAATAACCTT 241
QY 61 ttgaggggttaagaatacagggacatggtgttcacatttgctgcacagacacgcgc 120
DB 242 TCGGAGGGTTAGAAATGATGAGGACATGTTGTTCAATTCCTGCCACGAAACCGGC 301

QY 121 agcttcacttggaagacgaatcagccttggaagagatctccctaagcagagagaa 180
DB 302 AGCTTCACTTGGAACAGAAATCAGCGCTTGGAAGAGATCAATCCCTAAGCAGAGAGAA 361

QY 181 gctactaaag 190
DB 362 GCTACTAAAG 371

RESULT 9
LOCUS BG385575 1003 bp mRNA linear EST 12-MAR-2001
DEFINITION 602453872R1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4582504 5',
mRNA sequence.

ACCESSION BG385575
VERSION BG385575.1 GI:13278389
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS NIH-MGC http://img.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strusberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC

CNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov

Plate: LCM1304 row: o column: 17
High quality sequence stop: 694.

FEATURES

source Location/Qualifiers

1..1003
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4582504"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 282 a 248 c 264 g 209 t
ORIGIN

Query Match 99.2%; Score 188.4; DB 10; Length 1003;
Best Local Similarity 99.5%; Pred. No. 6.4e-47;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atattaagagagcctgaactgttccttgacatcttatgaatgacgaataactt 60
DB 73 ATATTAAAGAGAGCGCTGAACCTGTTCTTGACATCTTATTAATGTCAGAAATAACCTT 132
QY 61 ttgaggggttaagaatacagggacatggtgttcacatttgctgcacagacacgcgc 120
DB 133 TCGGAGGGTTAGAAATGATGAGGACATGTTGTTCAATTCCTGCCACGAAACCGGC 192

QY 121 agcttcacttggaagacgaatcagccttggaagagatctccctaagcagagagaa 180
DB 193 AGCTTCACTTGGAACAGAAATCAGCGCTTGGAAGAGATCAATCCCTAAGCAGAGAGAA 252

QY 181 gctactaaag 190
DB 253 GCTACTAAAG 262

RESULT 10
LOCUS AA315469 433 bp mRNA linear EST 19-APR-1997
DEFINITION EST187245 Colon carcinoma (HCC) cell line II Homo sapiens cDNA 5',
end, mRNA sequence.

ACCESSION AA315469
VERSION AA315469.1 GI:1967798
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult,
C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White,
O., Sutton, G., Blake, J.A., Brandon, R.C., Man, W.L., Clayton, R.A.,
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald,
L.M., Fitzhugh, W.M., Fritchman, J.L., Georgagen, N.S., Glodak, A.,
Guelm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Peligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y.,
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,

Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/cdb/hg1/hg1.html>)
Seq primer: M3 Reverse.
Location/Qualifiers
1..433
/organism="Homo sapiens"
/db_xref="ATCC (inhost):110433"
/db_xref="taxon:9606"
/clone_lib="Colon carcinoma (HCC) cell line II"
/tissue_type="colon"
/cell_line="KM12C(HCC)-parental human colon carcinoma ;Dukes B2"
/note="Organ: colon; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 112 a 101 c 103 g 112 t 5 others
ORIGIN
Query Match 98.9%; Score 188; DB 9; Length 433;
Best Local Similarity 98.9%; Pred. No. 6.4e-47;
Matches 188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 atattaaaggagagccgtgaacgttctctgtgacatctatgaatgcagaaataactct 60
Db 172 ATATTAAAGAGAGCCGGAACCTGTTCTTGACATCTTATGAATGCAAAAATACCTT 231
QY 61 ttggaagggttaagaagatcaagagcatggtgttcacattctgtccaggaacaccgc 120
Db 232 TTGGAGGGTTAGAAAGATCAGGAGCATGTTTTCACATTTGCTGCCACGAGACCCGCC 291
QY 121 agcttcacttgaagaacagatcacgcctgtgaaagatcatccttaagcagagagaa 180
Db 292 AGTTTTCACCTTGGAACAGATCAGCGCTTGTAAGAGATCACCCTAAGCAGAGAGAA 351
QY 181 gctactaaag 190
Db 352 GCTACTAAG 361
RESULT 11
BE868512 830 bp mRNA linear EST 20-OCT-2000
LOCUS 601444525F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3848358 5',
DEFINITION mRNA sequence.
ACCESSION BE868512
VERSION BE868512.1 GI:10317288
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 830)
NIH-MGC <http://mgi.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM9564 row: b column: 07
High quality sequence stop: 657.
Location/Qualifiers
1..830
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3848358"
/clone_lib="NIH_MGC_65"
/tissue_type="adenoecarcinoma"
/lab_host="PH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."
BASE COUNT 213 a 207 c 210 g 210 t
ORIGIN
Query Match 98.6%; Score 187.4; DB 10; Length 830;
Best Local Similarity 99.5%; Pred. No. 1.2e-46;
Matches 188; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 atattaaaggagagccgtgaacgttctctgtgacatctatgaatgcagaaataactct 60
Db 262 ATATTAAAGAGAGCCGGAACCTGTTCTTGACATCTTATGAATGCAAAAATACCTT 321
QY 61 ttggaagggttaagaagatcaagagcatggtgttcacattctgtccaggaacaccgc 120
Db 322 TTGGAGGGTTAGAAAGATCAGGAGCATGTTTTCACATTTGCTGCCACGAGACCCGCC 381
QY 121 agcttcacttgaagaacagatcacgcctgtgaaagatcatccttaagcagagagaa 180
Db 382 AGTTTTCACCTTGGAACAGATCAGCGCTTGTAAGAGATCACCCTAAGCAGAGAGAA 441
QY 181 gctactaa 189
Db 442 GCTACTCAA 450
RESULT 12
BE871910 912 bp mRNA linear EST 20-OCT-2000
LOCUS 601447936F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3851937 5',
DEFINITION mRNA sequence.
ACCESSION BE871910
VERSION BE871910.1 GI:10320686
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 912)
NIH-MGC <http://mgi.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LLM9573 row: g column: 10
High quality sequence stop: 624.
Location/Qualifiers

FEATURES

source

1..912

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3851937"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 262 a 227 c 234 g 189 t
ORIGIN

Query Match

Best Local Similarity 98.3%; Score 186.8; DB 10; Length 912;
Matches 188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atattaaagagagcctgaactgttccttgacatctatgaatgcagaaatacctt 60
|||||
Db 40 ATATTAAAGAGAGCCTGAACGTTCCTTGACATCTATGAAATGTCAGAAAATACCTT 99
|||||
QY 61 ttggagggttagaagatcagggacatgtgttcacattgtctgcacagacacggcc 120
|||||
Db 100 TTGGAGGGTTAAGATCAGGGGACATGTTGTTCACATTTCTGCGACGGAACCGGC 159
|||||
QY 121 agcttcacttggaacagacacgcctctgtgaagagatcaccctaaagcagagagaa 180
|||||
Db 160 AGCTTCACTTGGAACAGAAATCAGCCCTTGGAAGAGATCCTCCTAAGCAGAGAGAA 219
|||||
QY 181 gctactaaag 190
|||||
Db 220 GCTACTAAG 229

RESULT 13

BE617113 736 bp mRNA linear EST 20-OCT-2000
DEFINITION 601441649P1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3845677 5',
mRNA sequence.

ACCESSION BE617113
VERSION BE617113.1 GI:9888051
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9573 row: b column: 14
High quality sequence start: 9
High quality sequence stop: 670.

FEATURES

source

1..736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3845677"

/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 191 a 170 c 183 g 191 t 1 others
ORIGIN

Query Match

Best Local Similarity 93.7%; Score 178; DB 10; Length 736;
Matches 189; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 atattaaagagagcctgaactgttccttgacatctatgaatgcagaaatacctt 60
|||||
Db 161 ATATTAAAGAGAGCCTGAACGTTCCTTGACATCTATGAAATGTCAGAAAATACCTT 219
|||||
QY 61 ttggagggttagaagatcagggacatgtgttcacattgtctgcacagacacggcc 120
|||||
Db 220 TTGGAGGGTTAAGATCAGGGGACATGTTGTTCACATTTCTGCGACGGAACCGGC 279
|||||
QY 121 agcttcacttggaacagacacgcctctgtgaagagatcaccctaaagcagagagaa 180
|||||
Db 280 AGCTTCACTTGGAACAGAAATCAGCCCTTGGAAGAGATCCTCCTAAGCAGAGAGAA 339
|||||
QY 181 gctactaaag 190
|||||
Db 340 GCTACTAAG 349

RESULT 14

BI826675 763 bp mRNA linear EST 04-OCT-2001
DEFINITION 603077437P1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5169128 5',
mRNA sequence.

ACCESSION BI826675
VERSION BI826675.1 GI:15938225
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM11420 row: j column: 09
High quality sequence stop: 759.

FEATURES
source

1..763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5169128"
/clone_lib="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.3 kb, insert size range
0.9-3 kb. Library is normalized and enriched for

full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH-MGC library."

BASE COUNT 195 a 181 c 193 g 194 t

ORIGIN

Query Match 93.7%; Score 178; DB 10; Length 763;
Best Local Similarity 99.5%; Pred. No. 8.6e-44;
Matches 189; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 atattaaaggagagccctgaacatgcttccttggaacattatgaatgcagaaataacctt 60
|||||
Db 250 ATATTAAAGAGAGAGCCCTGAACCTGCTTCGACATCTTATGAATGTCAGAAAATACC-T 308
|||||
QY 61 ttggaagggttagaagatcagggagacatgctgttcacatttgcctgcaggaacacggcc 120
|||||
Db 309 TTGGAGGGTTAGAAAGATCAGGGGACATGTTGTTTCACATTGGCTGCCACGGAACCCGCC 368
|||||
QY 121 agcttcacttggaaacagaatcacgcctctgtgaagagatcccttaagcagagagaa 180
|||||
Db 369 AGTCTTCACCTTGAAGACAGATCACGCCCTTGTAAGAGATCATCCCTAAGCAGAGAGAA 428
|||||
QY 181 gctactaaag 190
|||||
Db 429 GCTACTTAAG 438

RESULT 15

BG747491

LOCUS 872 bp mRNA linear EST 15 MAY-2001
602704606F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4858104 5',
mRNA sequence.

ACCESSION

BG747491

VERSION

BG747491.1 GI:14058144

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 872)

AUTHORS

NIH-MGC http://mgc.nci.nih.gov/

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov

Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory

REFERENCE

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

AUTHORS

DNA Sequencing by: NIH Intramural Sequencing Center

TITLE

Clone distribution: MGC clone distribution information can be

JOURNAL

found through the I.M.A.G.E. Consortium/LLNL at:

COMMENT

http://image.llnl.gov

REFERENCE

plate: LNCM711 row: K column: 01

AUTHORS

High quality sequence stop: 747.

TITLE

Location/Qualifiers

JOURNAL

1. 872

COMMENT

/organism="Homo sapiens"

REFERENCE

/db_xref="taxon:9606"

AUTHORS

/clone="IMAGE:4858104"

TITLE

/clone_1b="NIH_MGC_15"

JOURNAL

/tissue_type="adenocarcinoma cell line"

COMMENT

/lab_host="DH10B (phage-resistant)"

REFERENCE

/note="Organ: colon; Vector: pOT87; Site_1: XhoI; Site_2:

AUTHORS

EcoRI; cDNA made by oligo-dT priming. Directionally

TITLE

cloned into EcoRI/XhoI sites using the following 5'

JOURNAL

adaptor: GGCACGAG(G). Size-selected >500bp for average

COMMENT

insert size 1.8kb. Library constructed by Ling Hong in

REFERENCE

the laboratory of Gerald M. Rubin (University of

AUTHORS

California, Berkeley) using ZAP-cDNA synthesis kit

TITLE

(Stratagene) and Superscript II RT (Life Technologies)"

JOURNAL

BASE COUNT 237 a 202 c 214 g 219 t

Query Match 93.7%; Score 178; DB 10; Length 872;
Best Local Similarity 99.5%; Pred. No. 9.1e-44;
Matches 189; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 atattaaaggagagccctgaacatgcttccttggaacattatgaatgcagaaataacctt 60
|||||
Db 62 ATATTAAAGAGAGAGCCCTGAACCTGCTTCGACATCTTATGAATGTCAGAAAATACC-T 120
|||||
QY 61 ttggaagggttagaagatcagggagacatgctgttcacatttgcctgcaggaacacggcc 120
|||||
Db 121 TTGGAGGGTTAGAAAGATCAGGGGACATGTTGTTTCACATTGGCTGCCACGGAACCCGCC 180
|||||
QY 121 agcttcacttggaaacagaatcacgcctctgtgaagagatcccttaagcagagagaa 180
|||||
Db 181 AGTCTTCACCTTGAAGACAGATCACGCCCTTGTAAGAGATCATCCCTAAGCAGAGAGAA 240
|||||
QY 181 gctactaaag 190
|||||
Db 241 GCTACTTAAG 250

Search completed: September 26, 2002, 02:54:56
Job time: 3821 sec

BASE COUNT

ORIGIN

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 04:22:05 ; Search time 4629.8 seconds
(without alignments)
858.793 Million cell updates/sec

Title: US-09-874-390-1_COPY_246_435
Perfect score: 190
Sequence: 1 atctaaagagcagcctgaa.....cagagagaagctactaaag 190

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
------------	-------------	-------	--------	----	-------------

1	190	100.0	2106	9	BC017032	BC017032 Homo sapi
2	190	100.0	2128	6	AR136078	AR136078 Sequence
3	190	100.0	2128	9	AF102542	AF102542 Homo sapi
4	190	100.0	187275	2	AC092755	AC092755 Homo sapi
5	188.4	99.2	2216	6	AX045249	AX045249 Sequence
6	188.4	99.2	2217	6	AX045251	AX045251 Sequence
7	188.4	99.2	2217	6	AX045253	AX045253 Sequence
8	37.4	19.7	169064	2	AC061965	AC061965 Homo sapi
9	36	18.9	150703	2	AC095344	AC095344 Rattus no
10	35.8	18.8	159098	2	AC016484	AC016484 Homo sapi
11	35.6	18.7	94742	2	AC006263	AC006263 Homo sapi
12	35.2	18.5	37731	2	HSB1D7	282173 Human DNA s
13	35.2	18.5	171902	2	AC066694	AC066694 Homo sapi
14	35	18.4	178367	2	AC006115	AC006115 Homo sapi
15	34.4	18.1	164368	2	AC073986	AC073986 Homo sapi
16	34.4	18.1	169386	2	AC053485	AC053485 Homo sapi
17	34.4	18.1	181663	2	AC092415	AC092415 Homo sapi
18	34.4	18.1	195280	2	AC025081	AC025081 Homo sapi
19	34	17.9	90922	2	AC094187	AC094187 Rattus no
20	34	17.9	165007	2	AC019047	AC019047 Homo sapi
21	33.2	17.5	167676	9	AL365356	AL365356 Human DNA
22	33	17.4	317	11	HSR055W65	251719 H. sapiens (
23	33	17.4	84300	9	HSU104A17	AL121861 Human DNA
24	33	17.4	172688	9	AL512641	AL512641 Human DNA
25	33	17.4	181570	2	AC068738	AC068738 Homo sapi
26	33	17.4	310721	2	AL627107	AL627107 Homo sapi
27	32.8	17.3	134149	2	AC013295	AC013295 Homo sapi
28	32.8	17.3	167116	9	AC009305	AC009305 Homo sapi
29	32.6	17.2	149914	9	AC073205	AC073205 Homo sapi
30	32.6	17.2	156306	2	HS873P14	AL031682 Human DNA
31	32.6	17.2	186436	2	AC068467	AC068467 Homo sapi
32	32.4	17.1	83167	9	AL590423	AL590423 Human DNA
33	32.4	17.1	110000	2	AC106860_1	Continuation (2 of
34	32.4	17.1	124000	2	AC004060	AC004060 Homo sapi
35	32.4	17.1	134354	2	AL645989	AL645989 Mus muscu
36	32.4	17.1	146851	2	AC027697	AC027697 Homo sapi
37	32.4	17.1	153936	2	AC012282	AC012282 Homo sapi
38	32.4	17.1	165606	2	AL512601	AL512601 Homo sapi
39	32.4	17.1	169638	9	AC009433	AC009433 Homo sapi
40	32.4	17.1	179000	2	AC011847	AC011847 Homo sapi
41	32.4	17.1	180915	8	GTAJ10592	AJD10592 Gulliarid
42	32.4	17.1	185854	2	AC027736	AC027736 Homo sapi
43	32.4	17.1	190925	2	AL606665	AL606665 Homo sapi
44	32.4	17.1	194103	2	AC060818	AC060818 Homo sapi
45	32.4	17.1	207296	2	AL645637	AL645637 Mus muscu

ALIGNMENTS

RESULT 1
LOCUS BC017032 2106 bp mRNA linear PRI 09-NOV-2001
DEFINITION Homo sapiens, glucosaminyl (N-acetyl) transferase 3, mucin type.
ACCESSION BC017032
VERSION BC017032.1 GI:16877561
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2106)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrived by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdpax11.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
 Series: IRAC Plate: 20 Row: n Column: 9
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 4758421.

FEATURES

source

1..2106
 /organism="Homo sapiens"
 /db_xref="LocusID:9245"
 /db_xref="taxon:9606"
 /clone="MGC:9086 IMAGE:3851937"
 /tissue_type="Colon, adenocarcinoma"
 /clone_id="N1H.MGC_65"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPO86"
 309..1625
 /codon_start=1
 /product="glucosaminyl (N-acetyl) transferase 3, mucin
 type"
 /protein_id="AAH17032.1"
 /db_xref="GI:16877562"
 /translation="MVQKRRLQQLHYLMAAGCYMLATVALKLSRLKCDSDHGLLES
 RESQSYCRNLIYNFLKLPARKSINSGVTRGDQDAVLQALINNLEVKKKRPFDTHT
 YLSLRDCEHFAERKFIQFPLSKSEVEPIASVYIHEKILNFERLLRAVYAPONTY
 CVHDEKSPETFEKAVKAIISCFPNVFTASKLVRYVYASWSVQADLNCMEDLLQSSV
 PKWYKLTGCTDEPIKSNEMVQALKMLNGRMSSEVPPKHEKTRMYTHFEVVDTL
 HLTNKKDPPVNTLMTGNAYIASRPNVQVLTENPKSQCLIEVVKQTYSPDEHLM
 TLQRRMRMGSPVNPHPKYDIDSMSTIARLVKMGHEGDIKCAPAPCSGIHQRAICV
 YGAGDNNMLQNHLLANKPEKVDNALQCLEELRYKAIYGTGL"

CDS

BASE COUNT 582 a 471 c 516 g 537 t
 ORIGIN

Query Match 100.0%; Score 190; DB 9; Length 2106;
 Best Local Similarity 100.0%; Pred. No. 5.1e-49;
 Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 aatataaagagagcctggaactgtctcttgacatctatgaatgtcagaataaccctt 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 59 ATATTAAAGAGGAGCCTGAAAGCTGTTCCCTTGACATCTTATGAAATGTCAGAAATACCTT 118
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 61 ttggaaggttagaagaatcagaggacatggttcttcacatttgctgcacaggaacaccgc 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 119 TTGGAGAGGTTAGAAAGATCAGGGGACATGTTGTTTCACATTTGCTGCGACGGAACACCGCC 178
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 121 agcttcacttggaacagaatcagccttctggaagaatcatccctaagcagagagaa 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 179 AGCTCTCAGCTTGGAACAAGATCAGCGCTTGTAAGAGATCATCCTTAAGCAGAGAGAA 238
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 181 gctactaaag 190
 ||||||||||||||||
 Db 239 GCTACTTAAG 248

RESULT 2
 LOCUS ARI36078 2128 bp DNA linear PAT 16-JUN-2001
 DEFINITION Sequence 1 from patent US 6136580.
 ACCESSION ARI36078
 VERSION ARI36078.1 GI:14476750
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 1
 1 Unclassified.

REFERENCE 1 (bases 1 to 2128)
 AUTHORS Fukuda, M. and Yeh, J.
 TITLE beta-1,6-N-acetylglucosaminyltransferase that forms core 2, core
 4 and I branches
 JOURNAL Patent: US 6136580-A 1 24-OCT-2000;
 FEATURES Location/Qualifiers
 source 1..2128
 /organism="unknown"

BASE COUNT 569 a 477 c 526 g 556 t
 ORIGIN

Query Match 100.0%; Score 190; DB 6; Length 2128;
 Best Local Similarity 100.0%; Pred. No. 5.1e-49;
 Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 aatataaagagagcctggaactgtctcttgacatctatgaatgtcagaataaccctt 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 104 ATATTAAAGAGGAGCCTGAAAGCTGTTCCCTTGACATCTTATGAAATGTCAGAAATACCTT 163
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 61 ttggaaggttagaagaatcagaggacatggttcttcacatttgctgcacaggaacaccgc 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 164 TTGGAGAGGTTAGAAAGATCAGGGGACATGTTGTTTCACATTTGCTGCGACGGAACACCGCC 223
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 121 agcttcacttggaacagaatcagccttctggaagaatcatccctaagcagagagaa 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 224 AGCTCTCAGCTTGGAACAAGATCAGCGCTTGTAAGAGATCATCCTTAAGCAGAGAGAA 283
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 181 gctactaaag 190
 ||||||||||||||||
 Db 284 GCTACTTAAG 293

RESULT 3
 LOCUS AF102542 2128 bp mRNA linear PRI 31-JAN-1999
 DEFINITION Homo sapiens beta-1,6-N-acetylglucosaminyltransferase mRNA,
 complete cds.
 ACCESSION AF102542
 VERSION AF102542.1 GI:4204683
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 2128)
 Yeh, J.C., Ong, E. and Fukuda, M.
 TITLE Molecular cloning and expression of a novel beta-1,
 6-N-acetylglucosaminyltransferase that forms core 2, core 4, and I
 branches
 JOURNAL J. Biol. Chem. 274 (5), 3215-3221 (1999)
 MEDLINE 99115671
 REFERENCE 2 (bases 1 to 2128)
 Yeh, J.C., Ong, E. and Fukuda, M.
 TITLE Direct Submision
 JOURNAL Submitted (28-OCT-1998) Glycobiology Program, The Burnham
 Institute, 10901 N. Torrey Pines Rd., La Jolla, CA 92037, USA
 FEATURES Location/Qualifiers
 source 1..2128
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="brain"
 /dev_stage="fetus"
 354..1670
 /note="glucosyltransferase; core 2 Gnt-M"

CDS

/codon_start=1
 /product="beta-1,6-N-acetylglucosaminyltransferase"
 /protein_id="AADI0824.1"
 /db_xref="GI:4204684"
 /translation="MVQKRRLQQLHYLMAAGCYMLATVALKLSRLKCDSDHGLLES
 RESQSYCRNLIYNFLKLPARKSINSGVTRGDQDAVLQALINNLEVKKKRPFDTHT
 YLSLRDCEHFAERKFIQFPLSKSEVEPIASVYIHEKILNFERLLRAVYAPONTY
 CVHDEKSPETFEKAVKAIISCFPNVFTASKLVRYVYASWSVQADLNCMEDLLQSSV

PMKYLNTCTGDEPIKSNAMVQALKMLNGNSMESEVPPKHETPMKHYHEVYRDL
HLNRKADPPYVLTMTGNAYIVASRDFVOHAKNPKSOOLIMVWDTYSPDEHMA
TLORARBMGSSVPHKPYDIDSMRTARIVKMOCHGSDIDKAPAYACSGIHORATCV
YGAIDLMMHLLANKRPYVDNALOCLEBYLTKATYGTETL

BASE COUNT 569 a 477 c 526 g 556 t

ORIGIN

Query Match 100.0%; Score 190; DB 9; Length 2128;
Best Local Similarity 100.0%; Pred. No. 5.1e-49;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atataaagagagcctgaacgtcttccttgacatctatgaatgctagaataacct 60
Db 104 ATATTAAAGAGAGCCTGAACGTCTTCTTGACATCTTATGAATGTCAGAAATACCTT 163
QY 61 ttgagaggttgaagatcagagagcattgttccatcttgctgcagcagacgcgc 120
Db 164 TTGAGAGGTTGAGAGATCAGGAGCATGTGTTCACATTTGCTGCCACGAGACACCCGC 223
QY 121 agtttcacttggaacagaaatcagccttgtagagatcatcccttaagcagagagaa 180
Db 224 AGTCTTCACTTGGAACAGAAATCAGCCTGTGGAAGAGATCATCCTTAAGCAGAGAGAA 283
QY 181 gctactaaag 190
Db 284 GCTACTAAG 293

RESULT 4
AC092755 187275 bp DNA linear HTG 19-JAN-2002
LOCUS Homo sapiens chromosome 15 clone RP11-361D15 map 15q22, ***
DEFINITION SEQUENCING IN PROGRESS ***, 2 ordered pieces.
AC092755 AC022480
AC092755.3 GI:18249989
VERSION HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM

REFERENCE 1 (bases 1 to 187275)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D., and Hood, L.

TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 187275)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D., and Hood, L.

TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT On Jan 19, 2002 this sequence version replaced gi:15022677.

----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMMSC
Web site: http://chroma.mbl.washington.edu/msg_www
Contact: leerowen@systemsbiology.org

----- Summary Statistics
Sequencing vector: pUC18, 108752
Chemistry: Dye-terminator Big Dye, 90% of reads
Chemistry: Dye-Primer Big Dye, 10% of reads
Assembly program: Phrap; version 0.990399
Insert size: -----; agarose-fp
Quality coverage: ---x in 920 bases; sum-of-coverage

----- Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 32680: contig of 32680 bp in length
* 32681 32780: gap of unknown length
* 32781 187275: contig of 154495 bp in length.
Location/Qualifiers
1. 187275
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q22"
/clone="RP11-361D15"
/clone_11b="RPC1 human BAC library 11"
/note="This clone overlaps RP11-112N19 and RP11-568G20"

BASE COUNT 53688 a 38780 c 38083 g 56624 t 100 others
ORIGIN

Query Match 100.0%; Score 190; DB 2; Length 187275;
Best Local Similarity 100.0%; Pred. No. 4.5e-49;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atataaagagagcctgaacgtcttccttgacatctatgaatgctagaataacct 60
Db 38571 ATATTAAAGAGAGCCTGAACGTCTTCTTGACATCTTATGAATGTCAGAAATACCTT 38630
QY 61 ttgagaggttgaagatcagagagcattgttccatcttgctgcagcagacgcgc 120
Db 38631 TTGAGAGGTTGAGAGATCAGGAGCATGTGTTCACATTTGCTGCCACGAGACACCCGC 38690
QY 121 agtttcacttggaacagaaatcagccttgtagagatcatcccttaagcagagagaa 180
Db 38691 AGTCTTCACTTGGAACAGAAATCAGCCTGTGGAAGAGATCATCCTTAAGCAGAGAGAA 38750
QY 181 gctactaaag 190
Db 38751 GCTACTAAG 38760

RESULT 5
AX045249 2216 bp DNA linear PAT 24-NOV-2000
LOCUS Sequence 1 from Patent WO0066727.
DEFINITION AX045249
ACCESSION AX045249
VERSION AX045249.1 GI:11343799
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM

REFERENCE 1 (bases 1 to 2216)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Adolf, G., Heider, K.H. and Sommergruber, W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 0066727-A 1 09-NOV-2000;
Boehringer Ingelheim International GmbH (BE)

FEATURES
source Location/Qualifiers
1. 2216
/organism="Homo sapiens"
/db_xref="taxon:9606"
5'UTR 1. 426

CDS

427..1743
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17336.1"
/db_xref="GI:11343800"
/translation="MYQKRLCOLHYLALGCTMLATYALKLSFRKCDSPHGLIES
RSOSQYCRNLIYNFLKLPKRSINCSGVTGRDQEDVLAIIINLEVRKRRPEFTDTH
YLSLRDEHFEKAEKRFIOEPLSKSEVEFPYAVSVIHEKIEFERLLRAVAPONTY
CVHDEKSEFEKAVKAIISCFPMVFIASKRVAVYASMSRVOADLNCMEDLQSSV
PMKYFLNCGTDFPIKSNAEVQALKMLGRMSSEVEPKHEKTRMKYHEFVVDTL
HLTKRKDPPEYVPMNFGNAYIASRDPVOHVILKRNKSOQLIEVYKDYSPDEHLMA
TIQRARMGSPVNPMPKYDISMTSIARLVKMOGEGDIDKCAPAPSCGTHQRAICV
YGADLNMMLQNHLLANKFDPKVDNDLQCLEEYLRKAIYGTGL"

3'UTR
BASE COUNT 600 a 499 c 547 g 570 t
ORIGIN

Query Match 99.2%; Score 188.4; DB 6; Length 2216;
Best Local Similarity 99.5%; Pred. No. 1.6e-48;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atataaagaagagcctggaactgtcttccttgagacatctatgaatgtcagaataatcctt 60
Db 177 ATATTAAAGAGGAGCCTGAAACTGTTCTTGACATCTTATGAAATGTCAGAAATACCTT 236

QY 61 ttggaaggttaagaagatcagggagacatggttgcacatttgcgcacaggaaccgcgc 120
Db 237 TTGAGAGGTTAGAGATCAGGGGACATGGTTGTTCAATTTGCTGCCACGAGACACCGCC 296

QY 121 agcttcacttggaacagaatcacgccttctgtgaagaagatcacccttaagcagaagagaa 180
Db 297 AGCTTCACCTGGGACAGATCAGCCCTTGGAAGAAGATCATCCCTAAGCAGAGAGAA 356

QY 181 gctactaaag 190
Db 357 GCTACTAAG 366

RESULT 6
LOCUS AX045251 2217 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 3 from Patent WO0066727.
ACCESSION AX045251
VERSION AX045251.1 GI:11343801
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2217)
AUTHORS Adolf.G., Heider.K.H. and Sommergruber.W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 0066727-A 3 09-NOV-2000;
Boehringer Ingelheim International GmbH (DE)
FEATURES
Location/Qualifiers
1..2217
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..844
5'UTR 1..426
CDS 427..999
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17335.1"
/db_xref="GI:11343802"
/translation="MYQKRLCOLHYLALGCTMLATYALKLSFRKCDSPHGLIES
RSOSQYCRNLIYNFLKLPKRSINCSGVTGRDQEDVLAIIINLEVRKRRPEFTDTH
YLSLRDEHFEKAEKRFIOEPLSKSEVEFPYAVSVIHEKIEFERLLRAVAPONTY
CVHDEKSEFEKAVKAIISCFPMVFIASKRVAVYASMSRVOADLNCMEDLQSSV
PMKYFLNCGTDFPIKSNAEVQALKMLGRMSSEVEPKHEKTRMKYHEFVVDTL
HLTKRKDPPEYVPMNFGNAYIASRDPVOHVILKRNKSOQLIEVYKDYSPDEHLMA
TIQRARMGSPVNPMPKYDISMTSIARLVKMOGEGDIDKCAPAPSCGTHQRAICV
YGADLNMMLQNHLLANKFDPKVDNDLQCLEEYLRKAIYGTGL"

3'UTR
BASE COUNT 600 a 499 c 547 g 570 t
ORIGIN

Query Match 99.2%; Score 188.4; DB 6; Length 2216;
Best Local Similarity 99.5%; Pred. No. 1.6e-48;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atataaagaagagcctggaactgtcttccttgagacatctatgaatgtcagaataatcctt 60
Db 177 ATATTAAAGAGGAGCCTGAAACTGTTCTTGACATCTTATGAAATGTCAGAAATACCTT 236

QY 61 ttggaaggttaagaagatcagggagacatggttgcacatttgcgcacaggaaccgcgc 120
Db 237 TTGAGAGGTTAGAGATCAGGGGACATGGTTGTTCAATTTGCTGCCACGAGACACCGCC 296

QY 121 agcttcacttggaacagaatcacgccttctgtgaagaagatcacccttaagcagaagagaa 180
Db 297 AGCTTCACCTGGGACAGATCAGCCCTTGGAAGAAGATCATCCCTAAGCAGAGAGAA 356

QY 181 gctactaaag 190
Db 357 GCTACTAAG 366

RESULT 7
LOCUS AX045253 2217 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 5 from Patent WO0066727.
ACCESSION AX045253
VERSION AX045253.1 GI:11343804
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2217)
AUTHORS Adolf.G., Heider.K.H. and Sommergruber.W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 0066727-A 5 09-NOV-2000;
Boehringer Ingelheim International GmbH (DE)
FEATURES
Location/Qualifiers
1..2217
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..844
5'UTR 1..426
CDS 427..999
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17337.1"
/db_xref="GI:11343805"
/translation="MYQKRLCOLHYLALGCTMLATYALKLSFRKCDSPHGLIES
RSOSQYCRNLIYNFLKLPKRSINCSGVTGRDQEDVLAIIINLEVRKRRPEFTDTH
YLSLRDEHFEKAEKRFIOEPLSKSEVEFPYAVSVIHEKIEFERLLRAVAPONTY
CVHDEKSEFEKAVKAIISCFPMVFIASKRVAVYASMSRVOADLNCMEDLQSSV
PMKYFLNCGTDFPIKSNAEVQALKMLGRMSSEVEPKHEKTRMKYHEFVVDTL
HLTKRKDPPEYVPMNFGNAYIASRDPVOHVILKRNKSOQLIEVYKDYSPDEHLMA
TIQRARMGSPVNPMPKYDISMTSIARLVKMOGEGDIDKCAPAPSCGTHQRAICV
YGADLNMMLQNHLLANKFDPKVDNDLQCLEEYLRKAIYGTGL"

3'UTR
BASE COUNT 600 a 499 c 547 g 572 t
ORIGIN

Query Match 99.2%; Score 188.4; DB 6; Length 2217;

CDS

427..1743
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17336.1"
/db_xref="GI:11343803"
/translation="MYQKRLCOLHYLALGCTMLATYALKLSFRKCDSPHGLIES
RSOSQYCRNLIYNFLKLPKRSINCSGVTGRDQEDVLAIIINLEVRKRRPEFTDTH
YLSLRDEHFEKAEKRFIOEPLSKSEVEFPYAVSVIHEKIEFERLLRAVAPONTY
CVHDEKSEFEKAVKAIISCFPMVFIASKRVAVYASMSRVOADLNCMEDLQSSV
PMKYFLNCGTDFPIKSNAEVQALKMLGRMSSEVEPKHEKTRMKYHEFVVDTL
HLTKRKDPPEYVPMNFGNAYIASRDPVOHVILKRNKSOQLIEVYKDYSPDEHLMA
TIQRARMGSPVNPMPKYDISMTSIARLVKMOGEGDIDKCAPAPSCGTHQRAICV
YGADLNMMLQNHLLANKFDPKVDNDLQCLEEYLRKAIYGTGL"

3'UTR
BASE COUNT 600 a 498 c 547 g 572 t
ORIGIN

Query Match 99.2%; Score 188.4; DB 6; Length 2217;
Best Local Similarity 99.5%; Pred. No. 1.6e-48;
Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atataaagaagagcctggaactgtcttccttgagacatctatgaatgtcagaataatcctt 60
Db 177 ATATTAAAGAGGAGCCTGAAACTGTTCTTGACATCTTATGAAATGTCAGAAATACCTT 236

QY 61 ttggaaggttaagaagatcagggagacatggttgcacatttgcgcacaggaaccgcgc 120
Db 237 TTGAGAGGTTAGAGATCAGGGGACATGGTTGTTCAATTTGCTGCCACGAGACACCGCC 296

QY 121 agcttcacttggaacagaatcacgccttctgtgaagaagatcacccttaagcagaagagaa 180
Db 297 AGCTTCACCTGGGACAGATCAGCCCTTGGAAGAAGATCATCCCTAAGCAGAGAGAA 356

QY 181 gctactaaag 190
Db 357 GCTACTAAG 366

RESULT 7
LOCUS AX045253 2217 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 5 from Patent WO0066727.
ACCESSION AX045253
VERSION AX045253.1 GI:11343804
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2217)
AUTHORS Adolf.G., Heider.K.H. and Sommergruber.W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 0066727-A 5 09-NOV-2000;
Boehringer Ingelheim International GmbH (DE)
FEATURES
Location/Qualifiers
1..2217
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..844
5'UTR 1..426
CDS 427..999
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17337.1"
/db_xref="GI:11343805"
/translation="MYQKRLCOLHYLALGCTMLATYALKLSFRKCDSPHGLIES
RSOSQYCRNLIYNFLKLPKRSINCSGVTGRDQEDVLAIIINLEVRKRRPEFTDTH
YLSLRDEHFEKAEKRFIOEPLSKSEVEFPYAVSVIHEKIEFERLLRAVAPONTY
CVHDEKSEFEKAVKAIISCFPMVFIASKRVAVYASMSRVOADLNCMEDLQSSV
PMKYFLNCGTDFPIKSNAEVQALKMLGRMSSEVEPKHEKTRMKYHEFVVDTL
HLTKRKDPPEYVPMNFGNAYIASRDPVOHVILKRNKSOQLIEVYKDYSPDEHLMA
TIQRARMGSPVNPMPKYDISMTSIARLVKMOGEGDIDKCAPAPSCGTHQRAICV
YGADLNMMLQNHLLANKFDPKVDNDLQCLEEYLRKAIYGTGL"

3'UTR
BASE COUNT 600 a 498 c 547 g 572 t
ORIGIN

Query Match 99.2%; Score 188.4; DB 6; Length 2217;

Oy	1	atataaagagagagccctaaactgttcccttggacattcaatgaatgctagaanaatracct	60
Db	177	ATATTAAGAGAGAGACCTGAACCTGTCTTGGACATCTTATGAATGTGAGAAATAATCTT	236
Oy	61	ttgaaggggttagaagatcaggggacatggtttgtcacattgtctgcacagacacgcgc	120
Db	237	TTGAGGGGTAGAGATCAGGGGACATGTTGTTTCACATTTGCTCTCCACGGAAACCGGC	296
Oy	121	agcttcactcttggaaacagatcacgccttcttgaagagatcatcccttaagcagagagaa	180
Db	297	AGCTTCTACTTGGGAAACGAATACGCGCTTGTGAAGAAATCATCCCTAAAGCAGAGAGAA	356
Oy	181	gctactaaag 190	
Db	357	GCTACTAAG 366	

RESULT	8
AC061965/c	
LOCUS	AC061965
DEFINITION	Homo sapiens chromosome 15 clone RP11-37J13 map 15, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.
ACCESSION	AC061965
VERSION	AC061965.12
KEYWORDS	GI:18643695
SOURCE	HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN. human.

ORGANISM
Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Primates: Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE
Homo sapiens chromosome 15, clone RP11-37J13
JOURNAL
Unpublished
2 (bases 1 to 169064)
AUTHORS
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Batowin, A., Barr, N., Bastien, V., Beada, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campioiano, A., Castle, A., Chepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Deatrelano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Glinde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karakas, A.,
 Klein, J., Larocque, K., Lamaszres, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquits, N.,
 McCarthy, M., McEwan, P., McGurt, A., McKenna, K., McPheters, R.,
 Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tessfaye, S., Theodore, J., Tirrell, A., Trevers, M., Trifillio, J.,
 Vassiliiev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

<http://ftp.genome.washington.edu/RM/Repeatasker.html>

 Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: <http://www.seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

 Project Information
 Center project name: L6995
 Center clone name: 37_J_13

```

*****
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 6616: contig of 6616 bp in length
* 6617 6716: gap of 100 bp
* 6717 169064: contig of 162348 bp in length.
*
location/Qualifiers
    1..169064

```

BASE COUNT	49937	a	37298	c	36780	g	44914	t	135	others
ORIGIN	/clone_11b-"RPCI-11 Human Male BAC"									

Query Match	19.7%	Score	37.4	DB	2	Length	169064
Best Local Similarity	55.9%	Pred. NO.	0.37				
Matches	71	Conservative	0	Mismatches	56	Indels	0
						Gaps	0

Oy	36	lctlttgatgattcgaanaataaccttittggaggttaagaagtcgcaggacbtgctgttc	95
Dd	80678	TTTTGAAATCCCGAATAATTAAGTGTGGTGAGTGTAGAAGAACTTGAAGCCTTTGTAC	80619
Oy	96	acaatttgtcgcaagaaacaccgccagctcttcactttgaaaagatatagaccttgtaa	155
Dd	80618	ACTCTGTGGTAGCCTTAATAATGGCACACCTGCTGTGAAAACATTATATACATTTTCTTAA	80559
Oy	156	gagatca	162
Dd	80558	ATCATTA	80552

RESULT	9
AC095344/c	
LOCUS	
DEFINITION	AC095344 150703 bp DNA linear HTG-20-DEC-2001
ACCESSION	Rattus norvegicus clone CH230-17J6, *** SEQUENCING IN PROGRESS ***,
VERSION	65 unordered pieces.
KEYWORDS	AC095344 AC095344.2 GI:17972951
SOURCE	HTG; HTGS_PHASE1.
ORGANISM	Norway rat.
	Rattus norvegicus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
	Rattus.
REFERENCE	1 (bases 1 to 150703)
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alstrooms, S. L., Amaizungu, H. C., Aye, J. R., Banks, T., Barbata, J., Benton, J., Bimege, K., Blankenburg, R., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N. P., Buhay, C., Burch, P., Burkett, C., Butrell, K. L., Byrd, N. C., Carron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R., Davila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O., Dem, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H., Dugan-Roches, S., Durbin, K. J., Earnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, U., Garcia, A., Garner, T., Garcia, N., Gill, R., Gorrell, J. H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Haws, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Homsl, F., Howard, S., Huber, J. J., Hulky, S., Hunne, J.,

```

TITLE      Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
REFERENCE  Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvas, J.,
AUTHORS    Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
TITLE      Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
JOURNAL     Louised, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
            Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
            Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,
            Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbal, K.,
            Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
            Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
            Ogulu, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
            Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
            Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,
            Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshart, N.,
            Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
            Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,
            Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
            Thomas, S., Usmani, K., Vasquez, L., Vera, Y., Villalon, D., Vinson, R.,
            Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
            Wallington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,
            Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
            Weinstein, G. and Gibbs, R.
            Direct Submission
            Unpublished
            2 (bases 1 to 150703)
            Direct Submission
            Worley, K.C.
            Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Dec 20, 2001 this sequence version replaced gi:15625898.
            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: GGYD
            Center clone name: CH230-11J6
            ----- Summary Statistics
            Assembly program: Phrap; version 0.990329first call to
            findPhrapList
            Consensus quality: 118800 bases at least Q40
            Consensus quality: 126709 bases at least Q30
            Consensus quality: 133759 bases at least Q20
            Estimated insert size: 121939; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-ef estimation
            Quality coverage: 1.9x in Q20 bases; sum-of-contigs estimation
            -----
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drafi_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 65 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            1
            8591: contig of 8591 bp in length
            * 8592      8691: gap of unknown length
            * 8692      14177: contig of 5486 bp in length
            * 14178      14277: gap of unknown length
            * 14278      18006: contig of 3729 bp in length
            * 18007      18106: gap of unknown length
            * 18107      21936: contig of 3830 bp in length
            * 21937      22036: gap of unknown length
            * 22037      27169: contig of 5133 bp in length
            * 27170      27269: gap of unknown length
            * 27270      30486: contig of 3217 bp in length
            * 30487      30586: gap of unknown length
            * 30587      33980: contig of 3394 bp in length
            * 33981      34080: gap of unknown length
            -----
            34081      37600: contig of 3520 bp in length
            * 37601      41166: gap of unknown length
            * 41167      41266: contig of 3466 bp in length
            * 41267      44820: gap of unknown length
            * 44821      44920: contig of 3554 bp in length
            * 44921      48562: gap of unknown length
            * 48563      48662: contig of 3642 bp in length
            * 48663      51497: gap of unknown length
            * 51498      51597: contig of 2835 bp in length
            * 51598      54539: gap of unknown length
            * 54540      54639: contig of 2942 bp in length
            * 54640      57599: gap of unknown length
            * 57599      57600: contig of 2960 bp in length
            * 57600      57699: gap of unknown length
            * 57700      61220: contig of 3521 bp in length
            * 61221      61320: gap of unknown length
            * 61321      63680: contig of 2360 bp in length
            * 63681      63780: gap of unknown length
            * 63781      66040: contig of 2260 bp in length
            * 66041      66147: gap of unknown length
            * 66141      68255: contig of 2115 bp in length
            * 68256      68355: gap of unknown length
            * 68356      71744: contig of 3389 bp in length
            * 71745      71845: gap of unknown length
            * 71845      74036: contig of 2191 bp in length
            * 74036      74136: gap of unknown length
            * 74136      75883: contig of 1748 bp in length
            * 75884      75983: gap of unknown length
            * 75984      78003: contig of 2020 bp in length
            * 78004      78103: gap of unknown length
            * 78104      80205: contig of 2102 bp in length
            * 80206      80305: gap of unknown length
            * 80306      82957: contig of 2651 bp in length
            * 82957      83057: gap of unknown length
            * 83057      85075: contig of 2019 bp in length
            * 85076      85175: gap of unknown length
            * 85176      87133: contig of 1958 bp in length
            * 87134      87233: gap of unknown length
            * 87234      89248: contig of 2015 bp in length
            * 89249      89348: gap of unknown length
            * 89349      91527: contig of 2178 bp in length
            * 91527      91627: gap of unknown length
            * 91627      93946: contig of 2320 bp in length
            * 93947      94046: gap of unknown length
            * 94047      95633: contig of 1587 bp in length
            * 95634      95733: gap of unknown length
            * 95734      97168: contig of 1435 bp in length
            * 97169      97268: gap of unknown length
            * 97269      98877: contig of 1609 bp in length
            * 98878      98977: gap of unknown length
            * 98978      101256: contig of 2279 bp in length
            * 101257      101356: gap of unknown length
            * 101357      102824: contig of 1468 bp in length
            * 102825      102924: gap of unknown length
            * 102925      104642: contig of 1718 bp in length
            * 104643      104742: gap of unknown length
            * 104743      105805: contig of 1063 bp in length
            * 105806      105905: gap of unknown length
            * 105906      107559: contig of 1654 bp in length
            * 107560      107659: gap of unknown length
            * 107660      109682: contig of 2023 bp in length
            * 109683      109782: gap of unknown length
            * 109783      111771: contig of 1989 bp in length
            * 111772      111871: gap of unknown length
            * 111872      113179: contig of 1308 bp in length
            * 113180      113279: gap of unknown length
            * 113280      115151: contig of 1872 bp in length
            * 115152      115251: gap of unknown length
            * 115252      116759: contig of 1508 bp in length
            * 116760      116859: gap of unknown length
            * 116860      117999: contig of 1140 bp in length
            * 118000      118099: gap of unknown length
            * 118100      119882: contig of 1783 bp in length

```

119883 119982: gap of unknown length
* 119983 121000: contig of 1018 bp in length
* 121001 121100: gap of unknown length
* 121101 122530: contig of 1430 bp in length
* 122531 122630: gap of unknown length
* 122631 124087: contig of 1457 bp in length
* 124088 124187: gap of unknown length
* 124188 125929: contig of 1742 bp in length
* 125930 126029: gap of unknown length
* 126030 127298: contig of 1269 bp in length
* 127299 127398: gap of unknown length
* 127399 128963: contig of 1565 bp in length
* 128964 129063: gap of unknown length
* 129064 130295: contig of 1332 bp in length
* 130296 130395: gap of unknown length
* 130396 131716: contig of 1321 bp in length
* 131717 131816: gap of unknown length
* 131817 133191: contig of 1375 bp in length
* 133192 133291: gap of unknown length
* 133292 134948: contig of 1657 bp in length
* 134949 135048: gap of unknown length
* 135049 136777: contig of 1728 bp in length
* 136777 136877: gap of unknown length
* 136877 137885: contig of 1009 bp in length
* 137886 137985: gap of unknown length

Query Match 18.9% Score 36; DB 2; Length 150703;
Best Local Similarity 60.0%; Pred. No. 1;
Matches 60; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Oy 55 tacccttggagggtgaagaatcaggagacatggtcttcacatttcctgcacgaac 114
DB 67223 TACGTTTGGAGTGTAGGACCTTGGGGATGCTGATTTTATTTTGTTCGGCCT 67164
Oy 115 accgcagcttccacttgcgaacagacacgccttgta 154
DB 67163 GGGCCAGTAGAGTCTTGAAACAGCTTCTCAGCTGTAA 67124

RESULT 10
AC016484 159098 bp DNA linear HTG 14-MAR-2001
LOCUS Homo sapiens chromosome 15 clone RP11-18H24, WORKING DRAFT
DEFINITION
SEQUENCE, 3 unordered pieces.
AC016484
AC016484.13 GI:13324773
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILL; HTGS_ACTIVEPIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 159098)
AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Mao,J., Komp,C., Kotler,S., Lam,B., Marathe,R., Miranda,M.,
Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
Southwick,A.M., Webb,C., Wilhelm,J., Yu,S. and Davis,R.W.
Unpublished
2 (bases 1 to 159098)
JOURNAL Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,
REFERENCE Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
AUTHORS Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D.,
Yu,S. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
COMMENT On Mar 14, 2001 this sequence version replaced gi:12331479.
----- Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center
Center site: http://sequence-www.stanford.edu/group/human/

Contact: hum-info@sequence.stanford.edu
----- Project Information
Center project name: 720
Center clone name: RP11-18H24
----- Summary Statistics
Sequencing Vector: M13mp18; X02513; 99% of reads
Sequencing Vector: plasmid; plasmid_accession: 1% of reads
Chemistry: Dye-primer; 1% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 157616 bases at least Q40
Consensus quality: 158170 bases at least Q30
Consensus quality: 158357 bases at least Q20
Insert size: 156454; agarose-fp
Insert size: 158898; sum-of-contigs
Quality coverage: 7.9x in Q20 bases; agarose-fp
Quality coverage: 8.2x in Q20 bases; sum-of-contigs.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2166: contig of 2166 bp in length
* 2167 2266: gap of unknown length
* 2267 44132: contig of 41866 bp in length
* 44133 44232: gap of unknown length
* 44233 159098: contig of 114866 bp in length.
Location/Qualifiers
1. 159098
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/clone="RP11-18H24"
/clone_11b="RPC1 human BAC library 11"
1. 2166
/note="assembly_name:Contig31"
2267. 44132
/note="assembly_name:Contig32"
44233. 159098
/note="assembly_name:Contig33
clone_end:SP6"

BASE COUNT 46075 a 33092 c 33649 g 46082 t 200 others
ORIGIN

Query Match 18.8% Score 35.8; DB 2; Length 159098;
Best Local Similarity 55.1%; Pred. No. 1.2;
Matches 70; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Oy 36 tctttagatgcaaatcatttggagggttagaagatcaggagacatggtgttc 95
DB 138922 TTTTGAAGAACCCAGAAATAAGTGTGTCGATGTAGAGAAAGTGAAGCCTGTAC 138981
Oy 96 acattgctgcacgcgaacacgcgcagcttccacttgcgaacagatcaagcctgtgaa 155
DB 138982 ACTCTGTGTGGAACATTAATAAGGACACAGCTGTGTGGAACATTAATGACATTTCCTTAA 139041
Oy 156 gagatca 162
DB 139042 ATCATTA 139048

RESULT 11
AC006263 94742 bp DNA linear PRI 01-JAN-1999
LOCUS Homo sapiens chromosome 17, clone hC17.187_K.10, complete sequence.
DEFINITION
AC006263
AC006263.1 GI:4092479
AC006263.1
KEYWORDS HTG.
SOURCE human.

ORGANISM	Homo sapiens	repeat_region	complement(4956..5273)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	repeat_region	/rpt_family="AluSx"
AUTHORS	1 (bases 1 to 94742)	repeat_region	complement(5316..5681)
TITLE	Homo sapiens chromosome 17, clone hC17.187_K.10	repeat_region	/rpt_family="MLT1B"
JOURNAL	Unpublished	repeat_region	5803..5889
REFERENCE	2 (bases 1 to 94742)	repeat_region	/rpt_family="L2"
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,D., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Herena,L., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.	repeat_region	6260..6553
TITLE	Submitted (30-DEC-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	repeat_region	/rpt_family="AluSg"
JOURNAL	Research, 320 Charles Street, Cambridge, MA 02141, USA	repeat_region	7723..7999
REFERENCE	3 (bases 1 to 94742)	repeat_region	/rpt_family="AluJo"
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,D., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Herena,L., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.	repeat_region	8000..8044
TITLE	Submitted (01-JAN-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	repeat_region	/rpt_family="(CA)n"
COMMENT	On Jan 1, 1999 this sequence version replaced g1:4079623. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html .	repeat_region	complement(8205..8506)
FEATURES	Location/Qualifiers	repeat_region	/rpt_family="AluSx"
source	1.94742	repeat_region	/rpt_family="MLT1J"
	/organism="Homo sapiens"	repeat_region	10430..10478
	/db_xref="taxon:9606"	repeat_region	/rpt_family="MER58A"
	/chromosome="17"	repeat_region	complement(10479..10776)
	/map="17"	repeat_region	/rpt_family="AluSg"
	/clone="hC17.187_K.10"	repeat_region	10777..10950
	/complement(11..109)	repeat_region	/rpt_family="MER58A"
	/rpt_family="MIR"	repeat_region	complement(11942..12248)
	complement(363..637)	repeat_region	/rpt_family="AluSx"
	/rpt_family="AluSx"	repeat_region	12505..12648
	643..689	repeat_region	/rpt_family="L2"
	/rpt_family="AT_rich"	repeat_region	16304..16379
	complement(1879..2173)	repeat_region	/rpt_family="MADE1"
	/rpt_family="AluSx"	repeat_region	complement(16984..17289)
	complement(2184..2496)	repeat_region	/rpt_family="AluSg"
	/rpt_family="AluX"	repeat_region	18258..18355
	2960..3003	repeat_region	/rpt_family="MIR"
	/rpt_family="AT_rich"	repeat_region	19744..19961
	complement(3438..3746)	repeat_region	/rpt_family="MIR"
	3817..4159	repeat_region	complement(20015..20289)
	/rpt_family="MER47A"	repeat_region	/rpt_family="L2"
	4867..4887	repeat_region	complement(22035..22411)
	/rpt_family="AT_rich"	repeat_region	/rpt_family="THE1C"
		repeat_region	complement(22492..22620)
		repeat_region	/rpt_family="MIR"
		repeat_region	complement(22621..22926)
		repeat_region	/rpt_family="AluSg"
		repeat_region	24191..24327
		repeat_region	/rpt_family="AluJo"
		repeat_region	24328..24635
		repeat_region	/rpt_family="AluSg"
		repeat_region	24636..24804
		repeat_region	/rpt_family="AluJo"
		repeat_region	complement(25514..25800)
		repeat_region	/rpt_family="AluSx"
		repeat_region	25829..26097
		repeat_region	/rpt_family="L2"
		repeat_region	26112..26211
		repeat_region	/rpt_family="L2"
		repeat_region	28376..28417
		repeat_region	/rpt_family="(CA)n"
		repeat_region	complement(28562..28598)
		repeat_region	/rpt_family="tRNA-Cys-TGC"
		repeat_region	complement(29242..29565)
		repeat_region	/rpt_family="AluSc"
		repeat_region	29686..29725
		repeat_region	/rpt_family="(CA)n"
		repeat_region	32426..32446
		repeat_region	/rpt_family="AT_rich"
		repeat_region	32450..32736
		repeat_region	/rpt_family="AluJo"
		repeat_region	32737..32766
		repeat_region	/rpt_family="(CAAA)n"
		repeat_region	complement(33301..33338)
		repeat_region	/rpt_family="A-rich"
		repeat_region	complement(34791..34833)
		repeat_region	/rpt_family="(TAA)n"
		repeat_region	36154..36312

```

repeat_region      /rpt_family="AluY"
                    36313..36339
                    /rpt_family="(CAAA)n"
                    complement(365756..37064)
repeat_region      /rpt_family="AluSx"
                    complement(37923..38269)
repeat_region      /rpt_family="THEIB"
                    39314..39753
                    /rpt_family="ML71C"
                    40962..41328
                    /rpt_family="THEIC"
                    complement(41915..42227)
repeat_region      /rpt_family="AluSx"
                    complement(42389..42489)
repeat_region      /rpt_family="MIR"
                    42542..42814
                    /rpt_family="Aluub"
                    43264..43294
repeat_region      /rpt_family="AT-rich"
                    44375..44680
repeat_region      /rpt_family="Aluuo"
                    46335..46413
                    /rpt_family="(CA)n"
                    complement(46830..47141)
repeat_region      /rpt_family="AluSg"
                    complement(47166..48452)
repeat_region      /rpt_family="LMC1"
                    complement(48835..49146)
repeat_region      /rpt_family="Aluuo"
                    50833..50862
                    /rpt_family="MIR"
                    complement(51012..51304)
repeat_region      /rpt_family="Aluuo"
                    51317..51358
                    /rpt_family="A-rich"
                    52808..52827
repeat_region      /rpt_family="(CAAA)n"
                    complement(53638..53974)
repeat_region      /rpt_family="AluSg"
                    54446..54480
repeat_region      /rpt_family="AT-rich"
                    55090..55110
                    /rpt_family="AT-rich"
                    complement(55411..55687)
repeat_region      /rpt_family="Aluuo"
                    complement(56518..56574)
repeat_region      /rpt_family="purine-rich"
                    57117..57403

```

Query Match 18.7% Score 35.6; DB 9; Length 94742;
 Best Local Similarity 54.6% Pred. No. 1.4;
 Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```

Oy 61 ttggaagggtgaagaatgaaggagatggtgttcatattgttcgacggaacacgccc 120
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 32641 TGGGAGATTCCTGAGCCGACGAGATGAGTGTAGTACGACGATACACCATCTGC 32700
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

Oy 121 agcttcacattggaacgaatcacgccttctgtaagagatcatccctaagcaagagaa 180
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 32701 ACTCCAGCCTGGCAGCAAGCAAGACCCCTATTCAAAAAACAAAACAAAACAAA 32760
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

Oy 181 gctactaaag 190
    ||  ||  ||  ||
Db 32761 GCAAAAAAAG 32770

```

RESULT 12 37731 bp DNA linear PRI 12-DEC-1999
 HSBID7 LOCUS
 DEFINITION Human DNA sequence from clone SC22CB-1D7 on chromosome 22 contains
 two exons of the LARGE gene for like-glycosyltransferase
 (K1AA0609), ESTs, an STS and GSSts, complete sequence.
 ACCESSION 282173

VERSION 282173.2 GI:6572307
 KEYWORDS HTG: glycosyltransferase; KIAA0609; LARGE.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 37731)
 AUTHORS Whiteley M.
 TITLE Direct Submission
 JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Requests: clonerequests@sanger.ac.uk
 On Dec 13, 1999 this sequence version replaced gi:3426102.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 22, constructed by the Sanger Centre Chromosome 22
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr22>
 SC22CB-1D7 is from the human chromosome 22-specific cosmid library
 (SC22CB) constructed at the Sanger Centre by Mark Ross and Cordelia
 Langford.

FEATURES
 source
 This sequence is the entire insert of clone SC22CB-1D7.
 VECTOR: lawr1516

Location/Qualifiers
 1..37731
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22"
 /clone="SC22CB-1D7"
 /clone_lib="SC22CB"
 1..77
 /note="AluY repeat: matches 2..78 of consensus"
 526..826
 /note="AluSg repeat: matches 1..310 of consensus"
 complement(608..1161)
 /note="match: GSS: Em:B18157"
 1144..1295
 /note="MIR repeat: matches 27..189 of consensus"
 1296..1591
 /note="match: STR: Em:G03780
 match: STR: Em:G03780"
 1305..1664
 /note="match: GSS: Em:B18152"
 1563..1634
 /note="L2 repeat: matches 2607..2666 of consensus"
 1645..1699
 /note="MIR1A1 repeat: matches 85..139 of consensus"
 complement(1657..2008)
 /note="match: GSS: Em:A0401622"
 1700..1802
 /note="AluSg/X repeat: matches 199..302 of consensus"
 2163..2558
 /note="match: GSS: Em:A0705678"
 2302..2413
 /note="MIR repeat: matches 29..153 of consensus"
 2497..2545
 /note="MIR1I repeat: matches 96..144 of consensus"

```

repeat_region 2687..2790 /note="MUT1H repeat: matches 428..534 of consensus"
repeat_region 2803..2886 /note="MIR repeat: matches 178..262 of consensus"
repeat_region 2961..3041 /note="L2 repeat: matches 2607..2675 of consensus"
repeat_region 3053..3105 /note="MUT1A1 repeat: matches 86..138 of consensus"
repeat_region 3106..3401 /note="AluY repeat: matches 1..295 of consensus"
repeat_region 3402..3614 /note="MUT1A1 repeat: matches 138..347 of consensus"
repeat_region 3622..3939 /note="L2 repeat: matches 2107..2438 of consensus"
repeat_region 3962..4144 /note="MER5A repeat: matches 1..189 of consensus"
repeat_region 4899..5194 /note="AluY repeat: matches 1..296 of consensus"
repeat_region 5457..5622 /note="MIR repeat: matches 37..214 of consensus"
repeat_region 5708..5940 /note="MIR repeat: matches 11..241 of consensus"
repeat_region 6008..6129 /note="MIR repeat: matches 100..214 of consensus"
repeat_region 6130..6395 /note="AluSc repeat: matches 34..301 of consensus"
repeat_region 6396..6487 /note="MIR repeat: matches 7..100 of consensus"
repeat_region 7028..7277 /note="L2 repeat: matches 2246..2497 of consensus"
misc_feature complement(7141..7684)
repeat_region 7521..7883 /note="MUT1E repeat: matches 180..564 of consensus"
repeat_region 7897..8003 /note="L2 repeat: matches 2583..2704 of consensus"
repeat_region 8459..8530 /note="L2 repeat: matches 2678..2748 of consensus"
repeat_region 8471..8565 /note="MIR repeat: matches 165..262 of consensus"
repeat_region 8868..9076 /note="AluY repeat: matches 6..214 of consensus"
repeat_region 9107..9269 /note="MER5B repeat: matches 1..172 of consensus"
repeat_region 9364..9862 /note="AluSg repeat: matches 1..299 of consensus"
gene complement(join(10020..10183,31581..31736))
CDS /gene="LARGE"
/note="match: CDNAS: Em:AJ006278 Em:AB011181 Em:AJ007583
match: ESTs: Em:AI630787 Em:T09171 Em:AA916781 Em:AA976610
Em:T31101 Em:H55069
match: proteins: Tr:O60348 Tr:O95461 Tr:Q921M7"
/codon_start=1
/evidence=not_experimental
/product="cbid7.1 (like-glycosyltransferase (KIAA0609))"
/protein_id="CAB62965.1"
/db_xref="GI:6572308"
/translation="VHWSPKRIKRVKKNHVEFRNLVLTPLFYDGNLLRFLFGPS
EADVNSENLQKQISELDDDLCTEFRRERFTVHRHLVFLHREYEPADSTDTLVAAQ
LSMD"
misc_feature complement(10020..10183)
/gene="LARGE"
/note="match: STS: Em:H55069"
repeat_region 10394..10701 /note="AluX repeat: matches 1..312 of consensus"
repeat_region 10867..11178 /note="AluX repeat: matches 1..312 of consensus"
repeat_region 11757..11875 /note="LIME3A repeat: matches 5993..6115 of consensus"
repeat_region 11890..12187 /note="AluYB repeat: matches 1..306 of consensus"

```

```

repeat_region 12248..12484 /note="LIP4 repeat: matches 5716..5952 of consensus"
repeat_region 12488..12684 /note="LIP4 repeat: matches 5948..6144 of consensus"
repeat_region 13025..13191 /note="MUT1H repeat: matches 25..204 of consensus"
repeat_region 13338..13630 /note="AluSg repeat: matches 1..291 of consensus"
repeat_region 14116..14381 /note="L2 repeat: matches 2298..2528 of consensus"
repeat_region 14922..15138 /note="MIR repeat: matches 8..261 of consensus"
repeat_region 15771..16053 /note="AluX repeat: matches 1..283 of consensus"
repeat_region 16776..17100 /note="AluY repeat: matches 1..311 of consensus"
repeat_region 17226..17283 /note="L2 repeat: matches 2648..2705 of consensus"
repeat_region 17235..17413 /note="MIR repeat: matches 48..254 of consensus"
repeat_region 17885..18193 /note="AluX repeat: matches 1..312 of consensus"
repeat_region 18197..18270 /note="37 copies 2 mer aa 67 conserved"
misc_feature complement(18347..19008)
/gene="LARGE"
/note="match: GSS: Em:AQ284466"
complement(18650..19035)
/gene="LARGE"
/note="match: GSS: Em:AQ167619"
repeat_region 18735..18772 /note="MIR repeat: matches 65..102 of consensus"
repeat_region 18832..19054 /note="MIR repeat: matches 15..240 of consensus"
misc_feature 19051..19425 /note="match: GSS: Em:AQ238543"
repeat_region 19218..19285 /note="MUT1H repeat: matches 103..172 of consensus"
repeat_region 19408..19480 /note="MUT1H repeat: matches 327..403 of consensus"
repeat_region 19529..19535 /note="could be 8 bp"
repeat_region 20201..20271 /note="L2 repeat: matches 2642..2710 of consensus"
repeat_region 20364..20600 /note="MIR repeat: matches 13..255 of consensus"
repeat_region 20651..20655 /note="could be 4T's"
repeat_region 21371..21503
Query Match 18.5%; Score 35.2; DB 9; Length 37731;
Best Local Similarity 49.5%; Pred. No. 1.9;
Matches 91; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY 6 aaagagagcctgaactctctcttgacatcttaagatgtaagaataacttttga 65
DB 10522 AAAAATTAGCCGACATGAGGTGTGCTGTAGTCTACTAGAGGCTGAGTGGGA 10581
QY 66 gggttagaagacatgaggaatggtgttcaacttgctgcacggaacccagctct 125
DB 10582 GAATTCACCTGAGCCCGGACATGAGGCTGCACTGAGCTGAGTGCACCTGC 10641
QY 126 tcacttgaacaacgaatcacgccttgtaagagatcatccctaagaagaagctac 185
DB 10642 AGCTAGGCTACAGACCAAAACCTGTTCAAAACAAACAAACAAACAAACAA 10701
QY 186 taata 189
DB 10702 AAAA 10705
RESULT 13
AC06694/c

```



```
repeat_region complement(3434, .3738)
/rpt_family="AluY"
repeat_region complement(3742, .3973)
/rpt_family="LIM4"
repeat_region 4031, .4343
/rpt_family="AluY"
repeat_region 4344, .4369
/rpt_family="AT-rich"
misc_feature 4578, .4762
/!note="predicted exon, program: graill2exons_human_1.3,
frame: 2, quality: good, score: 52.000"
repeat_region complement(4845, .4878)
/rpt_family="AT-rich"
repeat_region complement(4946, .5052)
/rpt_family="MER64"
repeat_region complement(5091, .5161)
/rpt_family="MER64B"
repeat_region 5163, .5461
/rpt_family="AluSc"
repeat_region complement(5699, .6017)
/rpt_family="MER64"
repeat_region 6396, .6492
/rpt_family="MIR"
repeat_region 6588, .6619
/rpt_family="AT-rich"
repeat_region complement(6655, .6956)
/rpt_family="AluSq"
misc_feature complement(6946, .7073)
/!note="predicted exon, program: graill2exons_human_1.3,
frame: 0, quality: excellent, score: 75.000"
repeat_region 7220, .7266
/rpt_family="LI"
repeat_region 7267, .7410
/rpt_family="BC200"
repeat_region 7411, .7464
/rpt_family="LI"
repeat_region 7467, .7763
/rpt_family="AluJo"
repeat_region 7764, .7860
/rpt_family="(GAAA)n"
repeat_region 7862, .8025
/rpt_family="LI"
repeat_region complement(8027, .8482)
/rpt_family="MER74"
repeat_region 8520, .8821
/rpt_family="AluSx"
repeat_region complement(8838, .9815)
/rpt_family="LFR5"
repeat_region complement(9873, .9972)
/rpt_family="(CA)n"
repeat_region complement(10307, .10723)
/rpt_family="MLT1B"
repeat_region complement(10935, .10967)
/rpt_family="(CA)n"
repeat_region complement(11309, .11400)
/rpt_family="(CA)n"
misc_feature complement(11693, .11874)
/!note="predicted exon, program: graill2exons_human_1.3,
frame: 2, quality: good, score: 59.000"
repeat_region 11966, .12160
/rpt_family="AluCb"
repeat_region 12272, .12292
/rpt_family="AT-rich"
repeat_region complement(12301, .12354)
/rpt_family="LIMB3"
repeat_region 12451, .12779
/rpt_family="MER7A"
repeat_region complement(12782, .13078)
/rpt_family="AluY"
repeat_region complement(13088, .13765)
/rpt_family="LIMB3"
repeat_region complement(13984, .14278)
/rpt_family="AluSx"
```

```
repeat_region 14710, .15013
/rpt_family="AluSx"
repeat_region 15016, .15039
/rpt_family="(CAA)n"
repeat_region complement(15740, .15809)
/rpt_family="MER64"
repeat_region 15906, .16202
/rpt_family="AluSx"
repeat_region complement(16482, .16563)
/rpt_family="MER74"
repeat_region complement(16787, .16902)
/rpt_family="(CA)n"
repeat_region complement(17091, .17426)
/rpt_family="MLT1B"
repeat_region complement(17590, .17630)
/rpt_family="(CA)n"
repeat_region 17882, .17942
/rpt_family="MIR"
misc_feature complement(17941, .18120)
/!note="predicted exon, program: graill2exons_human_1.3,
frame: 1, quality: excellent, score: 75.000"
repeat_region 18203, .18372
/rpt_family="AluSp"
repeat_region 18392, .18690
/rpt_family="AluSx"
repeat_region complement(18830, .18870)
/rpt_family="AT-rich"
repeat_region complement(18955, .18980)
/rpt_family="AT-rich"
repeat_region 19030, .19349
/rpt_family="MER7A"
repeat_region 19343, .19415
/rpt_family="(TA)n"
repeat_region complement(19417, .19701)
/rpt_family="AluSx"
repeat_region complement(19703, .19842)
/rpt_family="(TA)n"
repeat_region 19917, .20056
/rpt_family="(TA)n"
repeat_region complement(20066, .20210)
/rpt_family="AluSp/g"
repeat_region complement(20220, .20534)
/rpt_family="LIMB3"
repeat_region complement(20534, .20862)
/rpt_family="LIPAL6"
repeat_region complement(20863, .21162)
/rpt_family="AluSp"
repeat_region complement(21175, .21527)
/rpt_family="LIPAL6"
repeat_region complement(21574, .21847)
/rpt_family="AluY"
repeat_region complement(21848, .22139)
/rpt_family="LIMB3"
repeat_region 22346, .22366
/rpt_family="AT-rich"
repeat_region complement(22619, .22711)
/rpt_family="LIMB3"
repeat_region complement(23083, .23169)
/rpt_family="MER64B"
repeat_region 23170, .23296
/rpt_family="TIGER1"

Query Match 18,4% Score 35; DB 9; Length 178367;
Best Local Similarity 55,3% Pred. No. 2,1;
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 47 taagaataacttttgagggttagaatacaaggagacattggttcaacttgcctgc
DB 51135 TCAGGAACTGAGTGGAGGATCACTTGAGCTAGAGATGCGGTGACGTAAGCTGT
QY 107 caaggacacggccagcttcaacttgaacagagaatcaacgcttgaagagatcatccc
DB 51075 GATGGACCACTGCACTTCTGCTGGGCAACAGAGTAAGACCTGTGCTCAAAAAATCA 51016
```

OY 167 taa 169
 Db 51015 TAA 51013
 RESULT 15
 AC073986/c
 LOCUS
 DEFINITION Homo sapiens chromosome 5 clone RP11-427N4, WORKING DRAFT SEQUENCE,
 25 unordered pieces.
 AC073986
 VERSION AC073986.2 GI:9795960
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 164368)
 Waterston,R.H.
 The sequence of Homo sapiens clone
 Unpublished
 2 (bases 1 to 164368)
 Waterston,R.H.
 Direct Submission
 Submitted (08-JUL-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Aug 11, 2000 this sequence version replaced gi:8980003.
 COMMENT
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H_NH0427N04
 ----- Summary Statistics -----
 Sequencing vector: M13; 100%
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 150398 bases at least Q40
 Consensus quality: 154507 bases at least Q30
 Consensus quality: 156376 bases at least Q20
 Insert size: 16000; agarose-fp
 Insert size: 161968; sum-of-contigs
 Quality coverage: 3.75 in Q20 bases; sum-of-contigs
 Quality coverage: 3.82 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 1 1598: contig of 1598 bp in length
 * 1599 1698: gap of unknown length
 * 1699 3738: contig of 2040 bp in length
 * 3739 3838: gap of unknown length
 * 3839 4963: contig of 1125 bp in length
 * 4964 5063: gap of unknown length
 * 5064 7067: contig of 2004 bp in length
 * 7068 7167: gap of unknown length
 * 7168 9526: contig of 2359 bp in length
 * 9527 9626: gap of unknown length
 * 9627 12145: contig of 2519 bp in length
 * 12146 12245: gap of unknown length
 * 12246 15270: contig of 3025 bp in length
 * 15271 15370: gap of unknown length
 * 15371 17747: contig of 2377 bp in length

* 17748 17847: gap of unknown length
 * 17848 21615: contig of 3768 bp in length
 * 21616 21715: gap of unknown length
 * 21716 25342: contig of 3627 bp in length
 * 25343 25442: gap of unknown length
 * 25443 29469: contig of 4027 bp in length
 * 29470 29569: gap of unknown length
 * 29570 34443: contig of 4874 bp in length
 * 34444 34543: gap of unknown length
 * 34544 39595: contig of 5052 bp in length
 * 39596 39695: gap of unknown length
 * 39696 45834: contig of 6139 bp in length
 * 45835 45934: gap of unknown length
 * 45935 52225: contig of 6291 bp in length
 * 52226 52325: gap of unknown length
 * 52326 62588: contig of 10263 bp in length
 * 62589 62688: gap of unknown length
 * 62689 72677: contig of 9989 bp in length
 * 72678 72777: gap of unknown length
 * 72778 83043: contig of 10266 bp in length
 * 83044 83143: gap of unknown length
 * 83144 92619: contig of 9476 bp in length
 * 92620 92719: gap of unknown length
 * 92720 102437: contig of 9718 bp in length
 * 102438 102537: gap of unknown length
 * 102538 113400: contig of 10863 bp in length
 * 113401 113500: gap of unknown length
 * 113501 123821: contig of 10321 bp in length
 * 123822 123921: gap of unknown length
 * 123922 135948: contig of 12027 bp in length
 * 135949 136048: gap of unknown length
 * 136049 150259: contig of 14211 bp in length
 * 150260 150359: gap of unknown length
 * 150360 164368: contig of 14009 bp in length.
 Location/Qualifiers
 1. 164368
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="RP11-427N4"
 1. 1598
 /note="assembly_name:Contig5"
 1699. 3738
 /note="assembly_name:Contig6"
 3839. 4963
 /note="assembly_name:Contig7
 clone_end:T7
 vector_side:left"
 5064. 7067
 /note="assembly_name:Contig8"
 7168. 9526
 /note="assembly_name:Contig9"
 9627. 12145
 /note="assembly_name:Contig10"
 12246. 15270
 /note="assembly_name:Contig11"
 15371. 17747
 /note="assembly_name:Contig12"
 17848. 21615
 /note="assembly_name:Contig13"
 21716. 25342
 /note="assembly_name:Contig14"
 25443. 29469
 /note="assembly_name:Contig15"
 29570. 34443
 /note="assembly_name:Contig16"
 34544. 39595
 /note="assembly_name:Contig17"
 39696. 45834
 /note="assembly_name:Contig18"
 45935. 52225
 /note="assembly_name:Contig19"
 52326. 62588

[illegible]

Query Match	18.1%	Score 34.4	DB 2	Length 164368
Best Local Similarity	53.8%	Pred. No. 3.2		
Matches 71	Conservative 0	Mismatches 61	Indels 0	Gaps 0

Qy	45	tgctcgaaataacaccttttggagaggtatgaagaatccgagagacatggttctacatttct	104
Db	811/1	TCTCCATTTCCTCCCTTCCTTCAGGGGTGGGAGAGAATGTCAAGGGGTTCACATTMTA	811122
Qy	105	gcccagagacacacgcgcagctcttctacttggaaacagaatacaacgccttggagaagatc	164
Db	81111	GCAGCCAAAGCTGCTCCACCTCGGAGAGTGTAGAGAAAACCMATGATGATGATGAGAACAGA	810522
Qy	165	cctaagacagag	176
Db	81051	AAAAAGCATGAG	81040

Search completed: September 26, 2002, 04:23:02
Job time: 8947 sec

OS	Homo sapiens.
XX	
PN	CA2296936-A1.
XX	
PD	03-AUG-2000.
XX	
PF	03-FEB-2000; 2000CA-2296936
XX	
PR	03-FEB-1999; 99US-0118674
XX	
PA	(GLYC-) GLXCODESIGN INC.
XX	
PI	Korczak B, Lew A;
XX	

DR WPI: 2000-594746/57.
XX
XX New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
XX
PS Disclosure: Page 53; 66pp; English.
XX
XX The present sequence represents a noncoding region from a human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) gene.
CC The polypeptide can be used to treat diseases and disorders, including
CC cancer, cardiovascular disorders and inflammatory disorders, including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
SQ Sequence 267 BP; 66 A; 67 C; 60 G; 74 T; 0 other;

Query Match 100.0%; Score 190; DB 21; Length 267;
Best Local Similarity 100.0%; Pred. No. 1.6e-57;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atattaagaagagcctgaacactgtctccttgacacatcatatgaatgtcgaataactct 60
DB 18 atattaagaagagcctgaacactgtctccttgacacatcatatgaatgtcgaataactct 77

OY 61 ttggaagggttaagaagatcaggagacatggttcttcacatttgcgcacgaacaccgcc 120
DB 78 ttggaagggttaagaagatcaggagacatggttcttcacatttgcgcacgaacaccgcc 137

OY 121 agcttcacttggaaacagaatcacgccttctgaaagagatcaccttaagcagaagagaa 180
DB 138 agcttcacttggaaacagaatcacgccttctgaaagagatcaccttaagcagaagagaa 197

OY 181 gctactaaag 190
DB 198 gctactaaag 207

RESULT 2
ID AAA96570
XX AAA96570 standard; DNA; 2108 BP.
XX
AC AAA96570;
XX
DT 08-FEB-2001 (first entry)
XX
XX A core 2 beta-1,6-N-acetylglucosaminyltransferase DNA fragment.
DE
XX
XX Core 2 beta-1,6-N-acetylglucosaminyltransferase: core2b GlcNAc-T;
KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis; ss.
XX
XX Homo sapiens.
OS
XX CA2296936-A1.
PN
XX 03-AUG-2000.
PD

XX
XX 03-FEB-2000; 2000CA-2296936.
PF
XX
XX 03-FEB-1999; 99US-0118674.
PR
XX
XX (GLYC-) GLYCODESIGN INC.
PA
XX
XX Korczak B, Lew A;
PI
XX
XX WPI: 2000-594746/57.
DR
XX
XX
PT New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
XX
PS Claim 4; Page 51-52; 66pp; English.
XX
XX The present sequence encodes a partial human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
CC The polypeptide can be used to treat diseases and disorders, such as
CC cancer, cardiovascular disorders and inflammatory disorders, including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
SQ Sequence 2108 BP; 596 A; 464 C; 509 G; 539 T; 0 other;

Query Match 100.0%; Score 190; DB 21; Length 2108;
Best Local Similarity 100.0%; Pred. No. 4e-57;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atattaagaagagcctgaacactgtctccttgacacatcatatgaatgtcgaataactct 60
DB 18 atattaagaagagcctgaacactgtctccttgacacatcatatgaatgtcgaataactct 77

OY 61 ttggaagggttaagaagatcaggagacatggttcttcacatttgcgcacgaacaccgcc 120
DB 78 ttggaagggttaagaagatcaggagacatggttcttcacatttgcgcacgaacaccgcc 137

OY 121 agcttcacttggaaacagaatcacgccttctgaaagagatcaccttaagcagaagagaa 180
DB 138 agcttcacttggaaacagaatcacgccttctgaaagagatcaccttaagcagaagagaa 197

OY 181 gctactaaag 190
DB 198 gctactaaag 207

RESULT 3
ID AAC62134
XX AAC62134 standard; DNA; 2128 BP.
XX
AC AAC62134;
XX
DT 06-MAR-2001 (first entry)
XX
XX Nucleotide sequence of beta-1,6-N-acetylglucosaminyltransferase.
DE
XX
XX Human; beta-1,6-N-acetylglucosaminyltransferase; C2GNT-M; inflammation;
KW membrane protein; branched steryl lex; L-selectin; immune reaction;
KW inflammation; tissue rejection; tumour metastasis; ss.
XX

OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 354..1670
FT /tag= a
FT /product= "beta-1-6-N-acetylglucosaminyltransferase"
FT 2100..2105
FT /tag= b
XX
XX US6136580-A.
XX
XX 24-OCT-2000.
XX
XX 19-JAN-1999; 99US-0233506.
XX
XX 19-JAN-1999; 99US-0233506.
XX
XX (BURN-) BURNHAM INST.
XX
XX Fukuda M, Yeh J;
XX
XX WPI: 2001-040238/05.
XX P-PSDB; AAB30518.
XX
XX New C3Gnt-M polypeptides having core 2, core 4 and I branching
XX beta-1-6-N-acetylglucosaminyltransferase activities for preparing
XX reagents useful for diagnosing, preventing or treating inflammation or
XX tumour metastasis
XX
XX Example 1; Fig 4; 25pp; English.
XX
XX The present sequence encodes a human
XX beta-1-6-N-acetylglucosaminyltransferase polypeptide, with core2, core4
XX and I branching activities. It is designated C3Gnt-M. C3Gnt-M is a
XX membrane protein that is predominantly expressed in colon, small
XX intestine, trachea, stomach and thyroid, as well as in certain cancer
XX cell lines. C3Gnt-M polypeptides may be used to prepare molecules having
XX highly branched sialyl lex and I-selectins, which may be subsequently
XX used to modulate immune reactions, e.g. inflammation and tissue
XX rejection, and to prevent or inhibit tumour metastasis.
XX
XX Sequence 2128 BP; 569 A; 477 C; 526 G; 556 T; 0 other;
S0

Query Match 100.0%; Score 190; DB 22; Length 2128;
Best Local Similarity 100.0%; Pred. No. 4e-57;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atataaagagagcctgaacatgttctcttgacatcttatgaatgtaagaataactt 60
Db 104 atataaagagagcctgaacatgttctcttgacatcttatgaatgtaagaataactt 163
Qy 61 ttgaggggttgaagatcaaggagacatgttcttcaactttgctccacggaacccgc 120
Db 164 ttgaggggttgaagatcaaggagacatgttcttcaactttgctccacggaacccgc 223
Qy 121 agcttcacttggaacagaatcacgccttgtaagagatcatccctaagcagagagaa 180
Db 224 agcttcacttggaacagaatcacgccttgtaagagatcatccctaagcagagagaa 283
Qy 181 gctactaaag 190
Db 284 gctactaaag 293

RESULT 4
AAC99109
ID AAC99109 standard; cDNA; 2229 BP.
XX
XX AAC99109;
XX
XX 09-MAR-2001 (first entry)
XX

DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:337.
XX
XX Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
XX detection; diagnosis; identification; cytostatic; neuroprotective;
XX nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;
XX antiinflammatory; cardiant; gene therapy; chromosome mapping;
XX linkage analysis; tissue identification; tissue typing; forensic;
XX neural; immune system; muscular; reproductive; gastrointestinal;
XX pulmonary; cardiovascular; renal; proliferative; ss.
XX
XX Homo sapiens.
XX
XX WO200055320-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05989.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-579444/54.
XX P-PSDB; AAB54344.
XX
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
XX treating, or ameliorating a medical condition, particular pancreatic
XX cancer, or for use in assays for diagnosing a pathological condition -
XX
XX Claim 1; Page 759-760; 1379pp; English.
XX
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
XX proteins, called pancreatic cancer antigens, given in AAB54008 to
XX AAB54466. The human pancreatic cancer antigens have cytostatic,
XX neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,
XX gynaecological, cardiant and antiinflammatory activities, and can be used
XX in gene therapy. The polynucleotide and proteins can be used for
XX preventing, treating, or ameliorating a medical condition or in assays
XX for diagnosing a pathological condition or a susceptibility to one in a
XX subject. Binding partners to the proteins and the activity of the
XX proteins can be identified. The pancreatic cancer antigens can be used to
XX detect, treat or prevent pancreatic disorders, especially cancer.
XX Agonists and antagonists to the antigens can be screened for. The
XX pancreatic cancer antigen polynucleotides can be used to design nucleic
XX acid hybridisation probes that can be used in chromosome mapping, linkage
XX analysis, tissue identification and/or typing and a variety of forensic
XX and diagnostic methods. The proteins can be used to generate antibodies
XX which are used to purify, detect and target the polypeptides, including
XX both in vivo and in vitro diagnostic and therapeutic methods. The
XX proteins can be used to treat or prevent neural, immune system, muscular,
XX reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
XX proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
XX sequences used in the exemplification of the present invention.
XX
XX Sequence 2229 BP; 605 A; 501 C; 546 G; 572 T; 5 other;
S0

Query Match 100.0%; Score 190; DB 21; Length 2229;
Best Local Similarity 100.0%; Pred. No. 4.1e-57;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atataaagagagcctgaacatgttctcttgacatcttatgaatgtaagaataactt 60
Db 177 atataaagagagcctgaacatgttctcttgacatcttatgaatgtaagaataactt 236
Qy 61 ttgaggggttgaagatcaaggagacatgttcttcaactttgctccacggaacccgc 120
Db 237 ttgaggggttgaagatcaaggagacatgttcttcaactttgctccacggaacccgc 296
Qy 121 agcttcacttggaacagaatcacgccttgtaagagatcatccctaagcagagagaa 180

```

Db      297 agcttcacttggaaacagaatcagcgccttgtgaaagatcattcccttaagcaggagagaa 356
Qy      181 gctactaaag 190
        |||||||
Db      357 gctactaaag 366

RESULT  5
ID      AAH34463 standard; cDNA; 2236 BP.
XX
AC      AAH34463;
XX
DT      03-SEP-2001 (first entry)
XX
DE      Human colon cancer antigen encoding cDNA SEQ ID NO:1545.
XX
KW      Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW      colorectal carcinoma; chromosome 15; ss.
XX
OS      Homo sapiens.
XX
PN      WC200122920-A2.
XX
PD      05-APR-2001.
XX
PF      28-SEP-2000; 2000MO-US26524.
XX
PR      29-SEP-1999; 99US-0157137.
PR      03-NOV-1999; 99US-0163280.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX
PI      Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR      WPI: 2001-235357/24.
DR      P-PSDB; AAG75058.
XX
PT      Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT      useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS      Claim 1; Page 3190-3191; 9803pp; English.
XX
CC      AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC      cancer-associated nucleic acid molecules (N) and proteins (P), where
CC      the proteins are collectively known as colon cancer antigens. The colon
CC      cancer antigens have cytostatic activity and can be used in gene
CC      therapy and vaccine production. N and P may be used in the prevention,
CC      diagnosis and treatment of diseases associated with inappropriate P
CC      expression. For example, N and P may be used to treat disorders
CC      associated with decreased expression by rectifying mutations or deletions
CC      in a patient's genome that affect the activity of P by expressing
CC      inactive proteins or to supplement the patients own production of P.
CC      Additionally, N may be used to produce the colon cancer-associated PS,
CC      by inserting the nucleic acids into a host cell and culturing the cell
CC      to express the proteins. N and P can be used in the prevention, diagnosis
CC      and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC      and AAB77789 represent sequences used in the exemplification of the
CC      present invention.
CC      N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC      missing at time of publication, meaning no sequences are present for
CC      SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ      Sequence 2236 BP; 609 A; 503 C; 548 G; 574 T; 2 other;

Query Match      100.0%; Score 190; DB 22; Length 2236;
Best Local Similarity 100.0%; Pred. No. 4,1e-57;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      61  ttggaagggttagaagaatcaggggacatggtgtgtcacattgtgtccacggaacaccgcc 120
        |||||||
Db      244  ttggaagggttagaagaatcaggggacatggtgtgtcacattgtgtccacggaacaccgcc 303
Qy      121  agcttcacttggaaacagaatcagcgccttgtgaaagatcattcccttaagcaggagagaa 180
        |||||||
Db      304  agcttcacttggaaacagaatcagcgccttgtgaaagatcattcccttaagcaggagagaa 363
Qy      181  gctactaaag 190
        |||||||
Db      364  gctactaaag 373

RESULT  6
ID      AAA48623
XX
AC      AAA48623;
XX
DT      19-SEP-2000 (first entry)
XX
DE      Human C2/4GnT cDNA.
XX
KW      Human; C2/4GnT; UDP-N-acetylglucosamine; O-glycan biosynthesis;
KW      O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer; ss.
XX
OS      Homo sapiens.
XX
FH      Key      Location/Qualifiers
FT      CDS      496..1812
FT      /*tag= a
FT      /product= "C2/4GnT"
FT      primer_bind 586..605
FT      /*tag= b
FT      primer_bind 1794..1813
FT      /*tag= c
FT      polyA_signal 2244..2249
FT      /*tag= d
XX
PN      WO200034449-A2.
XX
PD      15-JUN-2000.
XX
PF      03-DEC-1999; 99WO-DK00677.
XX
PR      04-DEC-1998; 98DK-0001605.
XX
PA      (CLAU/) CLAUSEN H.
XX
PI      Clausen H, Schwientek T;
XX
DR      WPI: 2000-423407/36.
DR      P-PSDB; AAY94492.
XX
PT      New nucleic acid molecule encoding UDP-N-acetylglucosamine useful as
PT      probe for the detection of specified glucosaminyltransferase from
PT      other species and related organisms -
XX
PS      Claim 5; Fig 2; 47pp; English.
XX
CC      The present sequence encodes human UDP-N-acetylglucosamine:
CC      N-acetylglucosamine beta-1,6-N-acetylglucosaminyltransferase
CC      (C2/4GnT). The protein is the third member of the family of O-glycan
CC      beta-1,6-N-acetylglucosaminyltransferases. It adds core 2 or core 4
CC      based O-glycans on to oligosaccharides, glycoproteins and
CC      glycosphingolipids. C2/4GnT can therefore be used in the production of
CC      appropriately glycosylated glycoconjugates with particular enzymatic,
CC      immunogenic, or other biological or physical properties. The nucleotide
CC      sequence is useful as a probe for the detection of C2/4GnT from other
CC      species and related organisms and for the recombinant production of
CC      C2/4GnT polypeptide. The nucleotide sequence was identified by analysis
CC      of EST database sequence information. Oligonucleotides derived from EST

```


CC clone 17865 of ATCC were used to isolate two full-length C2/4Gnt clones
CC from a human foreskin genomic P1 library by 5' RACE PCR. RT-PCR was
CC performed using Coloz05 human cell line mRNA in order to produce cDNA
CC for expression of C2/4Gnt in Sf9. The control of O-glycan core assembly
CC has been implicated in tumour progression and metastasis.

XX Sequence 2319 BP; 615 A; 528 C; 559 G; 617 T; 0 other;

Query Match 100.0%; Score 190; DB 21; Length 2319;

Best Local Similarity 100.0%; Pred. No. 4.2e-57;

Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atataaagagagccttaaaccttctccttgacatctatcgaatgtaagaataactt 60
Db 246 atataaagagagccttaaaccttctccttgacatctatcgaatgtaagaataactt 305
QY 61 ttggaagggttagaagatcaggagacatggttgcacatttgcctgcacagacacgcgc 120
Db 306 ttggaagggttagaagatcaggagacatggttgcacatttgcctgcacagacacgcgc 365
QY 121 agcttcaccttggaacacgaatcacgccttgtaagagatcatccctaagcagagagaa 180
Db 366 agcttcaccttggaacacgaatcacgccttgtaagagatcatccctaagcagagagaa 425
QY 181 gctactaaag 190
Db 426 gctactaaag 435

RESULT 7

AAK94836
ID AAK94836 standard; cDNA; 2347 BP.

XX AAK94836;

XX 06-NOV-2001 (first entry)

XX Human full-length cDNA, SEQ ID NO: 3990.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.

XX EP1130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

XX 11-JAN-2000; 2000JP-0118774.

XX 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

XX Ota F, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI: 2001-524255/58.

XX P-PSDB: AAM93876.

XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

XX Claim 8; SEQ ID NO 3990; 1380bp + sequence listing; English.

XX The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full

CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 2347 BP; 555 A; 618 C; 751 G; 423 T; 0 other;

Query Match 16.7%; Score 31.8; DB 22; Length 2347;

Best Local Similarity 51.0%; Pred. No. 0.71;

Matches 75; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 6 aaagagagcctgaacacgttccttgagcattatgatgtcgaataatcctttga 65
Db 1867 aaagagagcctgaacacgttccttgagcattatgatgtcgaataatcctttga 1926
QY 66 gggttagaagatcagggagacatggtgtcacatttgcgcacagacacgcagctc 125
Db 1927 gggttagaagatcagggagacatggtgtcacatttgcgcacagacacgcagctc 1986
QY 126 tcaacttggaacagaatcacgccttgt 152
Db 1987 agcttggtcacagagtaagaccttgt 2013

RESULT 8

AAS69345
ID AAS69345 standard; cDNA; 2448 BP.

XX AAS69345;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #5149.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB: ABG05158.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 1; SEQ ID NO 5149; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or

PS Claim 1; SEQ ID No 8151; 103pp; English.

XX

XX

```

PR 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0486725.
PR 25-APR-2000; 2000US-0552317.
XX
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Dimaane RT;
XX WPI; 2001-457603/49.
DR P-PSDB; AAM25701.
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
PS Claim 1; Page 549-550; 1217pp; English.
XX
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antitumour;
CC antiallergic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virucide; anti-HIV; fungicide; antitumour;
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
CC antituber; osteopathic; dermatological; antiallergic; antisthmatic;
CC antidiabetic; cyostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
XX SQ Sequence 3652 BP; 1079 A; 796 C; 650 G; 1127 T; 0 other;

Query Match 16.4%; Score 31.2; DB 22; Length 3652;
Best Local Similarity 57.0%; Pred. No. 1.4;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 9 gagagagcctgaacgttctcttgagacatcttgaatgcagaaataacccttggagg 68
DB 2233 GAGGAGACAGAAAGTCATGTGCTATCTCTCGAAGATGACATAGACCTTTGGAGAG 2174
QY 69 ttgaagaatcagggagacatggtgttgcacattgctgcga 108
DB 2173 CGAGAGCATCAACAGGTGTTATCCAGAAAGTTGTTCCA 2134

RESULT 13
AAA09310
ID AAA09310 standard; DNA; 4422 BP.
XX
XX AAA09310;
AC
XX 10-AUG-2000 (first entry)
DT
XX Human cancer associated antigen precursor DNA, clone NY-BEN-45.
DE
XX
XX renal cancer; cancer associated antigen precursor; diagnosis;
KW
XX cyostatic; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH CDS 861..3308
FT /*tag=
FT /transl_except= (pos:1473..1480, aa:Ala)

```

```

FT
XX /transl_except= (pos:2214..2219, aa:Phe)
XX
XX WO200020587-A2.
PN
XX
XX 13-APR-2000.
PD
XX
XX 04-OCT-1999; 99WO-US22873.
PF
XX
XX 05-OCT-1998; 98US-0166300.
PR 05-OCT-1998; 98US-0166350.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
PA
XX Obata Y, Gout I, Tureci O, Sahin U, Pfeungschuh M, Scanlan MJ;
PI Stockert E, Chen Y, Old LJ, Jager E, Knuth A;
PI
XX WPI; 2000-303774/26.
DR P-PSDB; AAY92338.
XX
XX Preventing, diagnosing and/or treating disorders associated with
PT abnormal expression of human cancer associated antigens
PS
XX Claim 57; Page 80-81; 121pp; English.
XX
XX AAA09310-20 are novel genes isolated by SEREX screening from a renal
CC cancer cell line 1973/10.4. The genes encode cancer associated antigen
CC precursors. These gene products are useful in methods for preventing,
CC diagnosing and/or treating disorders, especially cancer, associated with
CC abnormal expression of human cancer associated antigens. The method
CC comprises contacting a sample from a subject with an agent that
CC specifically binds to the nucleic acid molecule or expression product
CC (or fragment) complexed with a human leukocyte antigen (HLA) molecule
CC and determining the interaction between the agent and the nucleic acid
CC molecule or the expression product as a determination of the disorder.
XX
XX SQ Sequence 4422 BP; 1355 A; 814 C; 978 G; 1270 T; 5 other;

Query Match 16.4%; Score 31.2; DB 21; Length 4422;
Best Local Similarity 57.0%; Pred. No. 1.5;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 9 gagagagcctgaacgttctcttgagacatcttgaatgcagaaataacccttggagg 68
DB 2280 gagagagacaaagatcagtagctatctctctggaatgacataggaccttggagg 2339
QY 69 ttgaagaatcagggagacatggtgttgcacattgctgcga 108
DB 2340 cgagacgatcaacaggtgtttatccagaaagtgttccca 2379

RESULT 14
AAS72731/C
ID AAS72731 standard; CDNA; 2936 BP.
XX
XX AAS72731;
AC
XX 13-FEB-2002 (first entry)
DT
XX DNA encoding novel human diagnostic protein #8535.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR

```

PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG08544.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
PS Claim 1; SEQ ID No 8535; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2936 BP; 935 A; 544 C; 664 G; 793 T; 0 other;

Query Match 16.1%; Score 30.6; DB 23; Length 2936;
Best Local Similarity 68.9%; Pred. No. 2.1;
Matches 42; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 2 tatlaagggagccgaactgttccttgacatctatgaatgcagaataactctt 61
Db 1892 TTTTTCAGAGACCCAGATCCCTGCTTGATATATATTATGTTGAATAACCTTC 1833
OY 62 t 62
Db 1832 t 1832

RESULT 15
AAC54273
ID AAC54273 standard; DNA; 2582 BP.
XX
AC AAC54273;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 77298.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX BP1033405-A2.
PN
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0133284.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135553.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.

```
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0161040.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 16.0%; Score 30.4; DB 21; Length 2582;
Best Local Similarity 59.1%; Pred. No. 2.3;
Matches 52; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 46 gtcagaataactcttgagggttgaagaatcaggagacatggtgttcacattgctg 105
      | ||||| || | || | || | || | || | || | || | || | || | || |
DB 1686 gacagaataactgtggaggaatcacagaatagatgactgtagtctaaagtagcgt 1745
      | || | || | || | || | || | || | || | || | || | || | || |

QY 106 ccaggaacacgcgcagcttcacttg 133
      | || | || | || | || | || | || | || | || | || | || | || |
DB 1746 ggtcttcacaacacacacttcttcg 1773
```

Search completed: September 26, 2002, 03:03:36
Job time: 4291 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 04:22:19 ; Search time 94.91 Seconds
(Without alignments)
491.733 Million cell updates/sec

Title: US-09-874-390-1_COPY_246_435

Perfect score: 190
Sequence: 1 atattaagaagagcgtctgaa.....cagagaagaactactaaag 190

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/1na/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/1na/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/1na/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/1na/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/1na/backt1es1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	190	100.0	2128 3 US-09-233-506-1	Sequence 1, Appli
2	27.2	14.3	720 6 5455030-14	Patent No. 5455030
3	27.2	14.3	84495 4 US-09-797-906-3	Sequence 3, Appli
4	27	14.2	4411529 4 US-09-103-840A-1	Sequence 1, Appli
5	26.4	13.9	13473 5 PCT-US96-03916-59	Sequence 59, Appli
6	26.4	13.9	18912 5 PCT-US96-03916-59	Sequence 21, Appli
7	26.2	13.8	2200 1 US-08-272-255-21	Sequence 21, Appli
8	26.2	13.8	2200 5 PCT-US95-08565-21	Sequence 44, Appli
9	26	13.7	1839 3 US-08-152-019A-44	Sequence 1, Appli
10	26	13.7	1839 3 US-08-482-677-1	Sequence 1, Appli
11	26	13.7	4895 3 US-09-053-866-1	Sequence 114, App
12	25.6	13.5	1730 2 US-08-687-080-114	Sequence 11, Appli
13	25.6	13.5	1948 4 US-09-360-197-11	Sequence 11, Appli
14	25.6	13.5	2688 1 US-08-088-656-3	Sequence 3, Appli
15	25.6	13.5	2688 1 US-08-245-756-3	Sequence 3, Appli
16	25.6	13.5	2688 1 US-08-441-750-3	Sequence 3, Appli
17	25.6	13.5	2688 2 US-08-441-751-3	Sequence 3, Appli
18	25.6	13.5	2688 5 PCT-US92-02521-3	Sequence 3, Appli
19	25.6	13.5	5893 2 US-08-687-080-44	Sequence 44, Appli
20	25.6	13.5	8982 3 US-08-976-255-5	Sequence 5, Appli
21	25.2	13.3	12537 2 US-08-611-280-4	Sequence 4, Appli
22	25.2	13.3	12537 4 US-09-195-940-4	Sequence 4, Appli
23	25	13.2	1676 4 US-09-143-954-3	Sequence 3, Appli
24	25	13.2	1702 4 US-09-143-954-3	Sequence 1, Appli
25	24.8	13.1	636 4 US-09-328-111-226	Sequence 226, App
26	24.8	13.1	969 4 US-08-936-165A-140	Sequence 140, App
27	24.8	13.1	5408 1 US-08-471-058-20	Sequence 20, Appli

28	24.8	13.1	5408 3 US-08-471-057-20	Sequence 20, Appli
29	24.6	12.9	1539 2 US-08-828-596-1	Sequence 1, Appli
30	24.6	12.9	1666 2 US-09-360-197-5	Sequence 5, Appli
31	24.6	12.9	1920 4 US-09-534-638-6	Sequence 6, Appli
32	24.6	12.9	2264 4 US-09-126-109-9	Sequence 9, Appli
33	24.6	12.9	2371 2 US-08-343-443B-1	Sequence 1, Appli
34	24.6	12.9	6156 4 US-08-891-640-1	Sequence 1, Appli
35	24.6	12.9	6156 4 US-09-723-535-3	Sequence 3, Appli
36	24.6	12.9	9840 4 US-09-534-638-1	Sequence 1, Appli
37	24.6	12.9	246240 2 US-08-724-394A-20	Sequence 20, Appli
38	24.6	12.9	246240 2 US-08-724-394A-21	Sequence 21, Appli
39	24.6	12.9	246240 2 US-08-724-394A-22	Sequence 22, Appli
40	24.4	12.8	243 4 US-08-905-223-63	Sequence 63, Appli
41	24.4	12.8	599 2 US-08-557-128-9	Sequence 9, Appli
42	24.4	12.8	2075 1 US-08-238-163-3	Sequence 3, Appli
43	24.4	12.8	35100 2 US-08-770-379-17	Sequence 17, Appli
44	24.4	12.8	35100 4 US-08-757-669A-17	Sequence 17, Appli
45	24.4	12.8	35100 4 US-09-230-371A-17	Sequence 17, Appli

ALIGNMENTS

RESULT 1
US-09-233-506-1
Sequence 1, Application US/09233506
Patent No. 6136580
GENERAL INFORMATION:
APPLICANT: Fukuda, Minoru
APPLICANT: Yeh, Jium Chern
TITLE OF INVENTION: A Beta-1-6-N-Acetylglucosaminyltransferase That Forms
FILE REFERENCE: P-LJ 3415
CURRENT APPLICATION NUMBER: US/09/233,506
CURRENT FILING DATE: 1999-01-19
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2128
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (354)..(1670)
US-09-233-506-1

Query Match 100.0%; Score 190; DB 3; Length 2128;
Best Local Similarity 100.0%; Pred. No. 6e-60;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atattaagaagagcgtctgaaactgtctcttgacatcttaagatgaagaataactt 60
|||||
Db 104 atattaagaagagcgtctgaaactgtctcttgacatcttaagatgaagaataactt 163

QY 61 ttgagaggttgaagaatacgaagagacatgttgcatacttgcctccaggaacacgcc 120
|||||
Db 164 ttgagaggttgaagaatacgaagagacatgttgcatacttgcctccaggaacacgcc 223

QY 121 agcttcacttggaacacgaatcagccttgcgaagatcatccctaagaagagaa 180
|||||
Db 224 agcttcacttggaacacgaatcagccttgcgaagatcatccctaagaagagaa 283

QY 181 gctactaaag 190
|||||
Db 284 gctactaaag 293

RESULT 2
5455030-14
Patent No. 5455030
APPLICANT: LADNER, ROBERT C.; BIRD, ROBERT E.; HARDMAN, KARL
TITLE OF INVENTION: IMMUNOTHERAPY USING SINGLE CHAIN

```
; POLYPEPTIDE BINDING MOLECULES
; NUMBER OF SEQUENCES: 24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/40,440
; FILING DATE: 1-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 512,910
; FILING DATE: 25-APR-1990
; APPLICATION NUMBER: 299,617
; FILING DATE: 19-JAN-1989
; APPLICATION NUMBER: 92,110
; FILING DATE: 02-SEP-1987
; APPLICATION NUMBER: 902,971
; FILING DATE: 01-SEP-1986
; SEQ ID NO:14
; LENGTH: 720
5455030-14

Query Match      14.3%; Score 27.2; DB 6; Length 720;
Best Local Similarity 58.8%; Pred. No. 1.7;
Matches 47; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 101 tctgtccacggaacacccgagctctcacttggaacagaatcacgccttgtaagagat 160
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 tgcctgcactattactgcacgacgacgacgacgacgacgacgacgacgacgacgac 311
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 161 catcccaacgacgagagaa 180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 312 caagcttgagctggaagata 331

RESULT 3
US-09-797-906-3/C
; Sequence 3, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION:
; APPLICANT: Zianghe YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: CLO01151CIP
; CURRENT APPLICATION NUMBER: US/09/797,906
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 84495
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(84495)
; OTHER INFORMATION: n = A,T,C or G
US-09-797-906-3

Query Match      14.3%; Score 27.2; DB 4; Length 84495;
Best Local Similarity 56.8%; Pred. No. 18;
Matches 50; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 65 aagggttagaagatcaggaggacatggtgttcacattgtctgcacggaacacccgacgtc 124
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31914 AGATTGCTTGTAACCTGGAGATGGAGGTTCAGATGAGGTGAGATGATGTCGCACTGCACATC 31855

QY 125 ttcaactggaacagaatcacgccttgt 152
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31854 TAGCCTGGCAACAGAGTGAAGACTTTGT 31827

RESULT 4
US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
```

```
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: ERASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      14.2%; Score 27; DB 4; Length 4411529;
Best Local Similarity 56.0%; Pred. No. 50;
Matches 51; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 8 agagagcctgaaactgttctcttgacatctatgaatgacgaataccttttgagg 67
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2404002 AGCAGCTGCGGAGCCTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2403943

QY 68 gttagaagatcaggaggacatggtgttcaca 98
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2403942 GCCTCAACGCTGCCGCATGTTGCGCAC 2403912

RESULT 5
PCR-US96-03916-1/C
; Sequence 1, Application PC/TUS9603916
; GENERAL INFORMATION:
; APPLICANT: Wild, Martha A.
; APPLICANT: Cochran, Mark D.
; TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCR/US96/03916
; FILING DATE: 23-MAR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/126,597
; FILING DATE: 24-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 39116-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13473 base pairs
; TYPE: nucleic acid
```

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1059..2489
FEATURE:
NAME/KEY: CDS
LOCATION: 2575..4107
FEATURE:
NAME/KEY: CDS
LOCATION: 4113..4445
FEATURE:
NAME/KEY: CDS
LOCATION: 4609..5487
FEATURE:
NAME/KEY: CDS
LOCATION: 5697..8654
FEATURE:
NAME/KEY: CDS
LOCATION: 9874..10962
FEATURE:
NAME/KEY: CDS
LOCATION: 11159..12658
FEATURE:
NAME/KEY: CDS
LOCATION: 12665..13447
PCT-US96-03916-1

Query Match 13.9%; Score 26.4; DB 5; Length 13473;
Best Local Similarity 61.8%; Pred. No. 14;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 6 aaagaggaagcctgaactgtctccttgagacatctatgaatgctcgaataatccttttga 65
Db 7923 ATAGGGGCGCTCGACGCTCGCTCGACATCTCATGATTCGTAACACAGTAGGAA 7864
QY 66 ggggttaga 73
Db 7863 ACGATAAA 7856

RESULT 6
PCT-US96-03916-59/C
Sequence 59, Application PC/TUS9603916
GENERAL INFORMATION:
APPLICANT: Wild, Martha A.
TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/03916
FILING DATE: 23-MAR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,597
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:

NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 39116-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 18912 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 697..1533
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: complement (1900..2784)
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: complement (2916..3605)
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 3694..5124
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 5210..7081
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 7245..8123
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 8333..11290
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 11098..12402
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 12510..13598
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 13792..15291
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 15298..16080
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 16129..17013
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: complement (17380..18216)
OTHER INFORMATION:
PCT-US96-03916-59

Query Match 13.9%; Score 26.4; DB 5; Length 18912;
Best Local Similarity 61.8%; Pred. No. 17;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY	6	aaaggggggcccgaacacgcttccctgggaacatttaagtcagaataataccctttga	65
Db	10559	ATAGGGGGGCTGCGACACTCTGCGCGACATCTCATGGAATTTCTGAAACACAGTAGGAA	105000
QY	66	gggttaga	73
Db	10499	ACGATGAA	10492

RESULT 7
 US-08-272-255-21/C
 Sequence 21, Application US/08272255
 Patent No. 5824859
 GENERAL INFORMATION:
 APPLICANT: Cashmore, Anthony R.
 APPLICANT: Ahmad, Margaret
 APPLICANT: Lin, Chentao
 TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
 TITLE OF INVENTION: Using the Same
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859/ris
 STREET: One Liberty Place, 46th floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/272,255
 FILING DATE: 08-JUL-1994
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: leary Ph.D., Kathryn
 REGISTRATION NUMBER: 36,317
 REFERENCE/DOCKET NUMBER: UPR-1795
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2200 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-272-255-21

```

RESULT      8
PCT-US95-08565-21/c
; Sequence 21, Application PC/TUS9508565
; GENERAL INFORMATION:
;   APPLICANT:  Cashmore, Anthony R.
;   APPLICANT:  Ahmed, Margaret

```

APPLICANT: Iln, Chertau
 TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
 TITLE OF INVENTION: Using the Same
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
 STREET: One Liberty Place, 46th floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/08565
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/272,255
 FILING DATE: 08-JUL-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Leary Ph.D., Kathryn
 REGISTRATION NUMBER: 36,317
 REFERENCE/DOCKET NUMBER: UPN-1795
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2200 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA

US-08-152-019A-44/c
Sequence 44, Application US/08152019A
Patent No. 5565331
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: Serfati, Tito
APPLICANT: Kennedy, Timothy
APPLICANT: Placzek, Marysia
APPLICANT: Jessell, Thomas
APPLICANT: Dodd, Jane
TITLE OF INVENTION: NEURAL AXON OUTGROWTH MODULATORS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEH, HOBNACH, TEST, ALBRITTON & HEBBERER
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

REFERENCE/DOCKET NUMBER: UC93-300-4

NAME/KEY:	Coding Sequence
;	


```

? FLING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 07/678,916
? FILING DATE: 01-APR-1991
? ATTORNEY/AGENT INFORMATION:
? NAME: Reiter, Stephen E
? REGISTRATION NUMBER: 31192
? REFERENCE/DOCKET NUMBER: 50848
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (619)552-1311
? TELEFAX: (619)552-0095
? TELEX: 20 6566 PATLAW CGO
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2688 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: unknown
? TOPOLOGY: unknown
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 643..1431
? FEATURE:
? NAME/KEY: mat_peptide
? LOCATION: 643..1431
? GS-08-088-633-3

```

[illegible]

RESULT 15
US-08-245-756-3/c
: Sequence 3, Application US/08245756
: Patent No. 5541112
: GENERAL INFORMATION:
: APPLICANT: Gleeson, Martin A
: APPLICANT: Howard, Bradley D
: TITLE OF INVENTION: GENES WHICH INFLUENCE PICHA PROTEOLYTIC
: TITLE OF INVENTION: ACTIVITY, AND USES THEREFOR
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
: STREET: 444 So. Flower Street, Suite 2000
: CITY: Los Angeles
: STATE: CA
: COUNTRY: USA
: ZIP: 90071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/245,756
: FILING DATE: 16-May-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/088,633
: FILING DATE: 06-JULY-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/678,916
: FILING DATE: 01-APR-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Seidman, Stephanie
: REGISTRATION NUMBER: 33,779

```

: . REFERENCE/DOCKET NUMBER: 9763
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-546-4737
: TELEFAX: 619-546-9392
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2688 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 643..1431
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 643..1431
:
: US-08-245-756-3

```

	Query Match	Similarity	66.1%	Score 25.6	DB 1	Length 2688
	Best Local	Similarity	66.1%	Pred. No. 12		
	Matches	37	Conservative	0	Mismatches 19	Indels 0
	Gaps	0				
Qy	20	aacttcttccttgagacatcttatagaatgtgcagaaaataccttlttgyaggttagaag	75			
b	1780	AGCCCTTAATACCCATTTTCTTAAGTTCACATTCATCATATTATTAGCAAGCTGGAAG	1725			

Search completed: September 26, 2002, 04:25:28
Job time: 9058 sec

• • • •

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 02:54:56 ; Search time 3532.1 Seconds
(without alignments)
7199.191 Million cell updates/sec

Title: US-09-874-390-1_COPY_436_2319
Perfect score: 1884
Sequence: 1 gattggtccctccaccact.....ataataaactcaacaga 1884

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	923.6	49.0	1862	11 AK008234	AK008234 Mus muscu
2	693.4	36.8	866	10 BG386598	BG386598 602454847
3	691.6	36.7	731	10 BG468641	BG468641 602510243
4	666	35.4	875	10 BG470202	BG470202 602533711
5	640	34.0	640	9 AA583339	AA583339 nm45f11.s
6	633.4	33.6	842	10 BE869192	BE869192 601445191
7	622.6	33.0	655	9 AM842622	AM842622 MR2-CN003
8	613.2	32.5	724	10 BG468448	BG468448 602532948
9	611	32.4	612	10 BE292814	BE292814 601105179
10	607.2	32.2	1050	10 BE867668	BE867668 601443206
11	605	32.1	671	9 A1587061	A1587061 tr55g02.x
12	602.6	32.0	766	10 B1765410	B1765410 603050445
13	599.8	31.8	603	9 AM192901	AM192901 x155h12.x
14	582.6	30.9	606	9 AM662377	AM662377 h125h05.x
15	561.2	29.8	864	10 BG821420	BG821420 602724876
16	557.2	29.6	823	10 BG169857	BG169857 602324322
17	552	29.3	659	10 BE304708	BE304708 601105907

18	518	27.5	872	10 BG747491	BG747491 602704606
19	516.4	27.4	554	9 AA307800	AA307800 EST178656
20	479.4	25.4	1003	10 BG385575	BG385575 602453872
21	463.2	24.6	609	10 BP881417	BP881417 PM0-ET020
22	445.8	23.7	826	10 BG821370	BG821370 602724814
23	436	23.1	436	9 AA566218	AA566218 n188g07.s
24	433	23.0	433	9 AA583146	AA583146 n37a11.s
25	416.2	22.1	920	10 BG386247	BG386247 602455439
26	396.6	21.1	912	10 BG469141	BG469141 602511125
27	394.6	20.9	912	10 BE871910	BE871910 601447936
28	383	20.3	383	9 A1150400	A1150400 qf40h07.x
29	381.2	20.2	567	12 A2388491	A2388491 IM0146G03
30	377.4	20.0	420	9 AM374998	AM374998 MR0-CT006
31	375.2	19.9	965	10 BF582709	BF582709 602094124
32	373.2	19.8	390	9 A1955582	A1955582 wt58e06.x
33	368.6	19.6	702	10 BG248304	BG248304 602400340
34	362.2	19.2	411	9 AM374874	AM374874 MR0-CT006
35	362.2	19.2	472	10 B1018802	B1018802 MR4-MT025
36	358	19.0	358	9 A1832775	A1832775 at63h05.x
37	353.6	18.8	629	10 BG167918	BG167918 602340003
38	344.2	18.3	357	10 BG984134	BG984134 I15-CN006
39	343.6	18.2	438	9 AM842590	AM842590 MR2-CN003
40	333.8	17.7	882	10 BG171085	BG171085 602324130
41	317.4	16.8	361	10 BF884849	BF884849 PM0-ET020
42	317	16.8	736	10 BE617113	BE617113 601441649
43	313.4	16.6	978	10 BG171515	BG171515 602322061
44	310.6	16.5	337	9 AM887761	AM887761 PM2-OT009
45	304.4	16.2	955	10 BG328280	BG328280 602427256

ALIGNMENTS

RESULT 1	AK008234	1862 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	Mus musculus adult male small intestine cDNA, RIKEN full-length				
DEFINITION	enriched library, clone:2010013H22:homolog to BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE, full insert sequence.				
ACCESSION	AK008234	GI:12842295			
VERSION	AK008234.1	GI:12842295			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2 (sites)				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20493374				
PUBMED	11042159				
REFERENCE	3 (sites)				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				


```

Db 1065 CGGCACTGATCGAGTGGGTAAGACACCTAATAGTCCCATGAGCACCTTTGGGACC 1124
Qy 1066 ctccagctgacagtgatgagctgctgctgtctcccaaccaccccaagtaacacatc 1125
Db 1125 CTCACAGCTGGCTGTGATGATGCTGATGATGATGATGATGATGATGATGATGATG 1184
Qy 1126 gacatgacttatttgcagagctgctgctgctgctgctgctgctgctgctgctgct 1185
Db 1185 GACATGAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1244
Qy 1186 ggtgctcttctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1245
Db 1245 GGGGACCTTACAGCTTCTTCTCAGAGATCCAGCCAGCGGCTGCTGCTTATGAGTCA 1304
Qy 1246 gggagactggaattgagctgctgctgctgctgctgctgctgctgctgctgctgctg 1305
Db 1305 GGGGACCTGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1364
Qy 1306 gtagatgataatgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1365
Db 1365 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1424
Qy 1366 actgactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1425
Db 1425 ACTGAACTATGAG--ACACTGTAAGACTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1477
Qy 1426 catgctgagactgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1485
Db 1478 -----CAGGACCTACTGAGGAGTGTGGGATGATGATGATGATGATGATGATG 1531
Qy 1486 tcccttgaataaagaagctgctgctgctgctgctgctgctgctgctgctgctgctg 1545
Db 1532 TCCCTTGAATGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1591
Qy 1546 ttgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1605
Db 1592 --GCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1642
Qy 1606 acttctcactaagtgagatgagactgctgctgctgctgctgctgctgctgctgct 1661
Db 1643 GCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1702
Qy 1662 atgtgtagagactgctgctgctgctgctgctgctgctgctgctgctgctgctg 1721
Db 1703 GTGACGAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1757
Qy 1722 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1781
Db 1758 ACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1817
Qy 1782 ccttccttctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1813
Db 1818 CCCCCCTTTTATATGTTAATTATTAACATPAA 1849

RESULT 2
LOCUS BG386598 866 bp mRNA linear EST 12-MAR-2001
DEFINITION 602454847P1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583243 5',
            mRNA sequence.
ACCESSION BG386598
VERSION BG386598.1 GI:13280044
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 866)
            NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-remail.nih.gov

```

```

Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.lnl.gov
Plate: L1CM1306 row: n column: 12
High quality sequence stop: 744.

FEATURES
    source
        1..866
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4583243"
            /clone_lib="NIH_MGC_15"
            /tissue_type="adenocarcinoma cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5'
            adaptor: GGACGAG(G). Size-selected >500bp for average
            insert size 1.8kb. Library constructed by Ling Hong in
            the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)".

BASE COUNT      219 a      209 c      221 g      217 t
ORIGIN
Query Match      36.8%; Score 693.4; DB 10; Length 866;
Best Local Similarity 97.9%; Pred. No. 1.2e-190;
Matches 755; Conservative 0; Mismatches 11; Indels 5; Gaps 5;

Qy 7 gtctctcccaacttccctgctgctgctgctgctgctgctgctgctgctgctgctgctg 66
Db 2 gttctctcccaacttccctgctgctgctgctgctgctgctgctgctgctgctgctgctg 61
Qy 67 caatggaagagactgctgcaactgctgctgctgctgctgctgctgctgctgctgctgctg 126
Db 62 CAATGGAAGAGACTGCTGCCAGCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 121
Qy 127 gccactgtgctctgnaacttcttccagtgtagtgtagtgtagtgtagtgtagtgtagt 186
Db 122 GCCACTGTGGCTGTGAAACTTCTTCAAGTTGAAGTGTGACTGTGACACCTGGGCTGTG 181
Qy 187 gattccaggaatctcaagccagctgtaggaatctgtataatcttccatgaactt 246
Db 182 GAGTCCA-GGAATTCAAAGCCAGTACGTAGGAATATCTGTAAATTTCTGAAACTT 240
Qy 247 ccagcaagaaggtctataactgttccaggtgtccaggaaggaaggaaggaaggaaggt 306
Db 241 CCAGCAAGAAGGTCTATACCTGTTCAAGGGTCCACCGAGGAGCAAGAGGAGGAGCTG 300
Qy 307 caggctattctgaataactctgagggtcaagaagaagcgaagccttccacagacacccac 366
Db 301 CAGGCTATTTCGAATTAACCTGAGGTCAAGAAGAAGAGAGGCTTTCACAGACACCCAC 360
Qy 367 taacttccctccacagagactgtgagcacttcaaggtcgaaggaaggtcatacagctc 426
Db 361 TACCTCTCCCTCACAGAGACTGTGACACTTCAAGGCTGAAGGTAAGTAATCATACGTT 420
Qy 427 ccactgagcaagaagaaggtgaggttccctattcattacattatgattatcatagaag 486
Db 421 CCACTGACCAAGAAGAGGTGAGTCCCTTTCATGATGATGATGATGATGATGATGATG 480
Qy 487 attgaaacttgaaggctactcagagctgtgatgcccctcagaacatactatgctc 546
Db 481 AFTGAAACTTTGAAGGCTACTGGAGCTGTGATGCCCTTCAGAACATATCTGCTG 540
Qy 547 catgtgatgagaagctcccaagaacttcaagaagcgtgtcaagaagcaatattctt-g 605
Db 541 CATGTGATGAGAGTCCCGCAAAACTTTCAAAGAGGCGGTCAAAAGCAATATTTCTTGG 600

```

QY 606 ctcccaaatgctctcaatgacgaagctggttcggtgttaactcctctcgt-cca 664
|||||
Db 601 CTTCCCAAAATGCTCTTCAAGGCTAAGCTGGTGGTGGTCTCTCTCTCTCCCA 660
QY 665 ggtgcaagctgacctaactgacatggaagactgtcccaagctca-gtgcggtgaa 723
|||||
Db 661 GGGTCAAGCTGACCTCAACTGACATGGAAGACTGTCTCCAGAGTCCCGGGGANA 720
QY 724 tactctcgtataacatgt-gggagcagactctcctataaagagcaatgcaga 773
|||||
Db 721 TACTTGTGAATTAATCATGTGGGGAGCGGCTTCTCTATAAGACATGGGAGA 771
RESULT 3
BG468641 731 bp mRNA linear EST 21-MAR-2001
LOCUS 602510243F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:464467 5',
DEFINITION mRNA sequence.
ACCESSION BG468641 GI:13400911
VERSION BG468641.1 GI:13400911
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 731)
NIH-MGC <http://mgs.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHCMI416 row: e column: 12
High quality sequence stop: 686.
location/Qualifiers
1. 731
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:464467"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOMB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using Zap-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT 214 a 157 c 183 g 177 t
ORIGIN
Query Match 36.7%; Score 691.6; DB 10; Length 731;
Best Local Similarity 98.4%; Pred. No. 3.6e-190;
Matches 720; Conservative 0; Mismatches 19; Indels 3; Gaps 2;

Db 122 AAGAGGCTATCAACTTTAGGGGTACCCGAGGGGACCAAGGAGGCTTCAAGCT 181
QY 313 attcgaataaccgtgaggtcgaagaagcgaagctctcaagaacccactacac 372
|||||
Db 182 ATTGTGAATTAACCTGGAGGCTCAGAGAGACGAGAGGCTTTCACAGACCCACTACTC 241
QY 373 tccctcacagagactgtgacacttcaagcgtgaaagaaagatcatacagttccactg 432
|||||
Db 242 TCCCTCACAGAGACTGTGAGCACTTCAAGGCTGAAAGAAATTCATACAGTCCCACTG 301
QY 433 agcaaaagaagagtggtggttccctattgcatctctatgtgtatctatgagaatgaa 492
|||||
Db 302 AGCAAAAGAGAGGTGGAGTTCCCTATTGCATCTATGATGATGATGATGATGAA 361
QY 493 aacttgaagagctacgtcagcgtgtgtatgccctcagaacatatactgtccatgtg 552
|||||
Db 362 AACTTTAAAGAGCTACGACGCTGTATGCCCTCAGAACTATATCTGTCTCAGTGTG 421
QY 553 gatgagaagtcgccagaacttccaaagaagcggtcaagaacattattctgtctccca 612
|||||
Db 422 GATGAGAGAGTCCCGAGAACTTCAAGAGGCGGTCAAGCAATTAATTTCTCTCCCA 481
QY 613 aatgtctcgaagcgaatgagctgtgtcgtgtgtatgcttccctcgtcaggtgcaa 672
|||||
Db 482 AATGCTTCAATAGCCACTAAGCTGTGCGGTGGTTATCCCTCTGCTCAGAGGTCAA 541
QY 673 gctgacctcaactgcaatgagaactgtctccagaagctcagtcggtgaaataactctg 732
|||||
Db 542 GCTGACTCAAGTGCATGAGAAAGACTTCTCCAGAGCTCATGCGGTGGAATCTTCTCTG 601
QY 733 aatacatgtggaagcagacttctcctataaagaacatgacagatgacagctctcaag 792
|||||
Db 602 AATTCATGTTGGTGGAGGAGCTTCTCTATTAAGAGCATGCAGATGATGTCAGG--CTCAAG 659
QY 793 atgtgaaatggaggaatgacatgaggt-cagaagtgacccctcaagacaagaagaacccg 851
|||||
Db 660 ATGTTGAATGGAGAGATACATGAGTACAGAGTACTCATTAAGCAAAAAGAAAAACG 719
QY 852 ctggaataatca 863
|||||
Db 720 TTGGAATATTA 731
RESULT 4
BG470202 875 bp mRNA linear EST 21-MAR-2001
LOCUS 602533711F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:461355 5',
DEFINITION mRNA sequence.
ACCESSION BG470202 GI:13402477
VERSION BG470202.1 GI:13402477
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 875)
NIH-MGC <http://mgs.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHCMI460 row: e column: 04
High quality sequence stop: 814.
location/Qualifiers
1. 875
FEATURES
Source

QY 1429 gctcagaacttgcctggagacagtgctgggtggagagccaggcttgcacattcgtgacacc 1488
Db 400 GCTCAGAACTTGTCTGGAGAGTGTGGGTGGAGACAGGGCTTGGCAATTCTGTGGCAATCC 341
QY 1489 tttagagtaagagggcgcctcattagattcgtggtaagtagatccttgccttgcaattg 1548
Db 340 TTTAGATTAAGAGGGCGCTGCTATTGTGATTTGGGTAAGTAGATCTTTTGCCCTTGCMAATTG 281
QY 1549 ctgcctgggtgtaatgctgcttctcctcaccaccacccacccacccacccacccacccaccc 1608
Db 280 CTGCCTGGGTAAGTGTCTGTCTCTCAACCCCTAACCCTAGTAGTCTCCCTCCACTAATC 221
QY 1609 ttccactatgtagaagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtaga 1668
Db 220 TTCTCCTAAGTAGAAGTAGAAGTAGAAGTAGAAGTAGAAGTAGAAGTAGAAGTAGAAGTAGAAGTAG 161
QY 1669 agagcactgattcactgaatgaatgcctgcctgtagcttccactcctgtagagctgcct 1728
Db 160 AGACCACTTGAATTCATTGATGCCCTGCTGGTATGCTTTCCATCTCTGTGAGCTGCCCT 101
QY 1729 tccataaatccaggttctgtagcgtggagaggaacttgaatggaagaagacccctccc 1788
Db 100 TCCATAAATTCCTCAGGTTGGTAGCCGTGAGAGAACTTGTGATGAAGAAGAACTTCCCTCC 41
QY 1789 ttctgactgttaactaaataaataaagctcctgattca 1828
Db 40 TTCTGTACTGTTAACCTTAATAATAATAGCTCTGATTTCA 1

RESULT 6
BE869192 842 bp mRNA linear EST 20-OCT-2000
LOCUS 601445191F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3849215 5',
DEFINITION mRNA sequence.
ACCESSION BE869192
VERSION BE869192.1 GI:10317968
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 842)
TITLE NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
plate: LLAM9566 row: e column: 24
High quality sequence stop: 684.
Location/Qualifiers
1. 842
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3849215"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: PCMV-SPORT6; Site: 1; Note:
Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

FEATURES
source
1. 842
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3849215"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: PCMV-SPORT6; Site: 1; Note:
Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 234 a 190 c 209 g 209 t
ORIGIN
Query Match 33.6%; Score 633.4; DB 10; Length 842;

Best Local Similarity 93.8%; Pred. No. 3.6e-173;
Matches 736; Conservative 0; Mismatches 41; Indels 8; Gaps 7;
QY 442 gaagtgaagtcctcattatgcatatcctatgtagtattcatatgagaagattgaaacttga 501
Db 1 GAGGTGAGAGTCCCTA-TGCATACTCATATGTGATTCATAGAGA-TGAATACTTTGAA 58
QY 502 aggcactgcagctgtgtatgcccctcagaacataactgtgtccatgtgtagaag 561
Db 59 AGGCTACTGCGAGCTGTGTATGCCCCCAGAAATATATAGTGTCCATCTGTGATGAGAAAG 118
QY 562 tcccccagaacttcataaagagcggtcaagcaattattctgtcttcccaatgtcttc 621
Db 119 TCCCCAGAAACTTTCAAAGAGCGGTAAAGCAATTAATTTCTTGCTTCCCAAAATGCTTC 178
QY 622 atagccagtaagctggtgtgtgttattatgctcctcgtgccaggggtgtcaagctgaccc 681
Db 179 ATAGCCAGTAGAGCTGCTGGGTGGTTATAGCCCTCGGTCCAGGGGTCAAGCTGACCTC 238
QY 682 aactgcatggaagactgtgtccagagctcagtgccgtgtgaataactctgtaatactgt 741
Db 239 AACTGCATGGAAGACTTGTCTCCAGAGCTCAGTCCCGTGGAAATACCTTCGTAATACATGT 298
QY 742 gggagcagcttctcctataaagagcaatgtagagatgtccaggctctcaagatgtgaat 801
Db 299 GGGAGCAGCTTCTCTATTAAGAGCAATGCAAGATGTGCTCCAGGCTCTCAAGATGTTGAAT 358
QY 802 gggagagatagcatgtagttagagtagtaccctcctaaagacaaagaacccgtgtgaatat 861
Db 359 GGGAGGAAATAGCATGAGTCAAGAGTACCTCTCAAGACAAAGAACCCCTGGAATAT 418
QY 862 caattgtaggttagtagagacacattacacctaaccacaagaagaagatcctccct 921
Db 419 CACTTGTAGGTAGTAGAGAGACACATTACCTTAACCAAGAAAGATTCCTCCCTCT 478
QY 922 tataatttaactatggtttaaagggaatggttaattgtgctccgaagtttcgtccaa 981
Db 479 TATTAATTAATGATGTTTAAAGGGAATGCGTAAATGTGCTTCCGAGATTTGCTCAA 538
QY 982 catg-ttttgaagaacccataaccacaacactgattgaatggtgtaaaagacac-ttata 1039
Db 539 CATGTTTGAAGAACCCCTTAATCCCAACAGATTTGAATGGGTAAACACTTTATA 598
QY 1040 gccacagatgaacacctctggtgccacccctcagcgtgtagcaggtgtgtagctgtgttc 1099
Db 599 GCCCAGATGAAGAACCTCTGTGGCCACCTTCAGCGTGACCGTGAATGCCCTG-TCTGTTC 657
QY 1100 ccaaccaccccaagtagacatctcagacatgacatgacatgacatgacatgacatgacatgac 1159
Db 658 CCAAC-AACCAAGTAGACATCTCAGACATGATTTCTATTGCGCGGTGATTAATGGC 715
QY 1160 aggtcatgagggagagacatcgatgaaggtgtccttattgctcctgtcctgtaaccacc 1219
Db 716 CGGTCAATGAGGAGCAATTCGATTAAGGCGTGTATGCTCTTGT-TCTGGAATCCAC 774
QY 1220 agcgg 1224
Db 775 ACGGG 779
RESULT 7
AM842622/c 655 bp mRNA linear EST 18-MAY-2000
LOCUS AM842622
DEFINITION MR2-CN0035-220300-204-f09 CN0035 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM842622
VERSION AM842622.1 GI:7936605
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 655)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,

TITLE
JOURNAL
MEDLINE
COMMENT

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-MR2-CN0035-220300-204-f09&f3=2000-03-22&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 652.

FEATURES

source

1. .655

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CN0035"
/dev_stage="Adult"
/note="Organ: colon,normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT

151 a 162 c 152 g 190 t

ORIGIN

Query Match 33.0%; Score 622.6; DB 9; Length 655;
Best Local Similarity 98.6%; Pred. No. 4.4e-170;
Matches 628; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

275 gggtcaccgaggagcccaagaagcagtgctcagctctcgaataacttgaggtca 334
|||||
637 GGGTCACCCGAGGGGACCAAGAGGCGAGTCTCAAGCTATTCTGAATACCTGAGGTC 578
|||||
335 agaagaagcgaagccttcacagacacccactcctccctcacagagacttgagc 394
|||||
577 AGAGAAACGAGAGCGCTTTCACAGACACCACTACTCTCCCTCACAGACACTGTGAGC 518
|||||
395 actccaagctgaagaagatcacaagttcccaactgagcaagaagagtgagattcc 454
|||||
517 ACTTCAGAGCGTGAAGGAGTTCATACAGTCCCACTGAGCAAGAAAGAGTGAGTCC 458
|||||
455 ctattgactctatggttgatctatgagaagattgaaacttgaaagtgctactgag 514
|||||
457 CTATTGCACTCTATGCTGATTCATGAGAGATTGGAAGCTTTGAAAGGCTACTGCGAG 398
|||||
515 ctgtgtatgcccccaagaacatactatgcatgtgagagagaagccccagaactt 574
|||||
397 CTGTGTATGCCCTCAGACACATATACCTGTGCATGTGATGAGAGAGGCCACAGAACTT 338
|||||
575 tcaagaagggggtcaagaactatcttctgctcccaaatgctctcactagaccagtaagc 634
|||||
337 TCAAGAAGAGCGGTCCAGCAATTAATTTCTCTCTCCCAATGATGCTCTATACCCAGTAAAG 278
|||||
635 tggttggagtggttatgagccctggttcagaggtgcaagcgcgactcaactgcatgagag 694
|||||
277 TGGTTGGGgtttatgctcctcctggttcagaggtgcaagcgcgactcaactgcatgagag 218
|||||

695 actgtcccaagagctcagtcgctggaataactctctgaatacatgtggagcggacttc 754
|||||
217 ACTGCTCCAGAGCGTCATGCGCGTGAATATCTTCTGAATACATGTGGAGCGACTTTC 158
|||||
755 ctataaagacgaatgcaagatggttcagcagctctcaagaatttgaaatggaggaataga 814
|||||
157 CTATTAAGAGCAATGCACAGATGTGTCAGGCTCTCAAGATGTGAATGGAGGATACGA 98
|||||
815 tggagtcagagtgctctctcctaagacaaagaccgctggaaatataacttgagtag 874
|||||
97 TGAAGTCAGAGTACTCTCTTAAGCACAAGAAACCCGCTGGAATATCATTGTGAGTAG 38
|||||
875 tgaagacacattcacctaaccaacaaagaagaaga 911
|||||
37 TGAGAGACACATTACCTTACCAACAATATACGA 1
|||||

RESULT 8
BG469448 724 bp mRNA linear EST 21-MAR-2001
LOCUS
DEFINITION 602532948F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4660835 5',
mRNA sequence.
ACCESSION BG469448
VERSION BG469448.1 GI:13401723
KEYWORDS EST.
SOURCE human.
ORGANISM human.

REFERENCE 1 (bases 1 to 724)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: LNCM458 row: O column: 12
High quality sequence stop: 687.

FEATURES

source

1. .724

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4660835"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF8; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 180 a 176 c 179 g 188 t

ORIGIN

Query Match 32.5%; Score 613.2; DB 10; Length 724;
Best Local Similarity 95.9%; Pred. No. 2.5e-167;
Matches 694; Conservative 0; Mismatches 23; Indels 7; Gaps 6;

7 gtctctccacgtctcctctgtg-ctcggtccacacgtgtcccatctctgagagatggt 65
|||||
2 gtctctctccacgtctcctctgtg-ctcggtccacacgtgtcccatctctgagagatggt 61
|||||
66 tcaatggaagagactctgcagcgtcatcttctgtggctctggtgctatctatgctgct 125
|||||

```

Db      62 TCATGGAAGAGACTGCGAGCTGCAATTCCTTGCGCTTGCTGATATGCTGCT 121
      126 ggcacatctggtctgaacattctcttcagggtgaagtgaactgcagaccactgggtc 185
      122 GGCACATGTGGCTCTGAACCTTCTTTCAGGTGGAAGTGTGACTGCACACTGGGTCT 181
      186 ggaagtcagggaatctcaagccagactcgttagaataatctgtataattccatgaact 245
      182 GGAGTCCA-GGAATCTCAAGAGCCAGTACTGTAGGAATATCTTATATATTTCTGAAACT 240
      246 tccaagaagaaggtctataactggttcagggtgcaccgaaggagaccagaagagcagtgct 305
      241 TCAGGAAGAAGGCTATCACTGTTCAAGGGGTCACCCGAGGGGACCAAGAGGACAGTCT 300
      306 tcaagctattctaaataacactggaagtcagaagaagcagacccttcacaacaccca 365
      301 TCAGGCTATTCTGAATACCTGAGAGTCAGAGAAAGCCAGAGCCTTTTCACAGACACCA 360
      366 ctactctccctccaccagagactgtgagcaactcaaggctgaaggaggtcatalactagt 425
      361 CTACTCTCCCTCCACAGAGACTGTGAGCACTTCAAGGCTGAAGAAGATTCATACAGTT 420
      426 cccactgagcaagaagaaggtgaggtccctatgcatatctctatgtgtaataaga 485
      421 CCACGTGAGCAAGAAGAGGTGAGTCCCTATTGATATCTATGCTGATGATGATGAGA 480
      486 gattgaaacttgaagactgagctgagctgtgtatgccccacagaa-catalactgtg 544
      481 GATTGAAACTTTGAAAGGCTACTGCGAGCTGTGTATGCCCTCAGAACATATCTACTGTC 540
      545 tccatgtgagtgaagaagtcacca--gaacttcaagaagggcgtaaagaactatctc 602
      541 TCCATGTGATGAGAGTCCCAAGAAAGATTTTCAAGAGGCGGTC-AAGCAATTTATTC 599
      603 ttgctcccaaatgcttcaatagccag-taagctgtgtcgggtgtgttctgctcctgtg 661
      600 TTGCTCCCAATGCTTCTACAGCTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659
      662 ccaaggtgcaagctgacactcaactgcatggaagaagctccagacacacagtcagtcg 721
      660 TCCAGGTGCAAGCTGACCAATATGATGGAGCATGGAGCATTTGCTACAGAGCTGCTGCT 719
      722 aata 725
      720 AATA 723

RESULT 9
BE292814 612 bp mRNA linear EST 13-JUL-2000
LOCUS BE292814 601105179P1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:2988090 5',
DEFINITION mRNA sequence.
ACCESSION BE292814
VERSION BE292814.1 GI:9175472
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 612)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM76 row: m column: 19

```

```

FEATURES             High quality sequence stop: 605.
source               Location/Qualifiers
                    1..612
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:2988090"
                    /clone_1ib="NIH_MGC_15"
                    /tissue_type="adenocarcinoma cell line"
                    /lab_host="DH10B (phage-resistant)"
                    /note="Organ: colon; Vector: pOTB7; Site: 1: XhoI; Site-2:
                    EcoRI; cDNA made by oligo-dt priming. Directionally
                    cloned into EcoRI/XhoI sites using the following 5'
                    adaptor: GGCACAG(G). Size selected >500bp for average
                    insert size 1.8kb. Library constructed by Ling Hong in
                    the laboratory of Gerald M. Rubin (University of
                    California, Berkeley) using ZAP-cDNA synthesis kit
                    (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT          181 a 137 c 140 g 154 t
ORIGIN
Query Match          32.4%; Score 611; DB 10; Length 612;
Best Local Similarity 100.0%; Pred. No. 1e-166;
Matches 611; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
      455 ctatgcatactctatggtgatcatgagaagattgaaacttgaagagctactgcag 514
      1 CTAATGCATACCTGATGATGATCATGAGAAATGAAAACCTTGAAGGCTACTCGAG 60
      515 ctgtgtatgccccctcagaacataactgtgtcgaatgtgataagaagtcaccagaact 574
      61 CTGTGTATGCCCTCGAAGCATATGATGTCATATGTGATGATGAGAGTCCCAAGAAACT 120
      575 tcaagaagcggtcagaagaactatctctgtctcccaaatgtctcatagccaagaagc 634
      121 TCAGAAAGGCGGTCAAGAGCAATATTCTTGTCTCCCAATATGCTCATAGCGAGTAAGC 180
      635 tgggtcgggtgttatgctctccctcgttccaagggtgcaagctgacactcaactgataga 694
      181 TGGTTCGGGTGTTTATGCTCTCTGCTCCAGGTCGCAACCTGCATCAACTGCATGGAAG 240
      695 actgtccagaagctgaagtcggtggaatactctctgaatacatgttggaagagacttc 754
      241 ACTGCTCCAGAGCTCAGTGGCCGTGGAATATCTTCTGATATCATGTGGAGCGACTTTC 300
      755 ctataagaagcaatgcagatagtgctcaggctctccaagatgttgaatgggaagataaga 814
      301 CTATAAAGAGCAATGCAGAGATGCTCAGGCTCTCAAGATGTTGAATGGAGAGATAGCA 360
      815 tggagtcagaagtaactctcttaagcacaagaagaaccgcttggaataatacttgagtag 874
      361 TGGAGTCAGAGGTACCTCTCTAAGCACAAAGAAACCCGCTGGAAATATCACTTGAAGTAG 420
      875 tgaagaacacattacaccctcaacaacaagaagaagatctccccccttaattatira 934
      421 TGAGAGACACATTACCTTACCTACCAAGAAAGAGATCTCCCTTATTAATTAACTA 480
      935 tgtttacaggaatgcgtacatgtgtgtctccgaagattcgtccacatgttttgaaga 994
      481 TGTTTACAGGAATGCGTACATTGTGTGCTCCGAAATTTGCGCCAACTGTTTGAAGA 540
      995 acccttaatccacaacaactgattgaatgggtlaaagaacctatagccagatgaacacc 1054
      541 ACCCTTAATCCCAACAACGATGATGAGGTAAAGACACTTAATAGCCAGATGAACACC 600
      1055 tctggggcacc 1065
      601 TCTGGGCCACC 611

RESULT 10
BE867668 1050 bp mRNA linear EST 20-OCT-2000
LOCUS BE867668

```


DEFINITION	601443206c1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847441 5', mRNA sequence.			
ACCESSION	BE667668			
VERSION	BE667668.1 GI:10316444			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 1050)			
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .			
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1AM9561 row: 1 column: 02 High quality sequence stop: 599. Location/Qualifiers			
FEATURES	1..1050			
SOURCE	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3847441" /clone_1lb="NIH_MGC_65" /tissue_type="adenocarcinoma" /lab_host="DH10B (phage resistant)" /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dN. Average insert size 1.8 kb. Library constructed by Life Technologies."			
BASE COUNT	390 a	242 c	231 g	187 t
ORIGIN				
Query Match	32.2%;	Score 607.2;	DB 10;	Length 1050;
Best Local Similarity	86.9%;	Pred. No. 1.8e-165;		
Matches 702; Conservative	0;	Mismatches 103;	Indels 3;	Gaps 3
OY	357	agacaccactactctccctcaccagagactgtgacactcaagctgaagaaagt	416	
Db	1	AGACACCCCACTACTCTCCCTCACCGAGACTGTGAGCACTTCAAGCTGAAGGAGATT	60	
OY	417	catacaattcccaactgagagaagaagaagttgagttccctattgtacatactatgtgat	476	
Db	61	CATTACAGTGTCCACATGAGCAAAAGAAAGGTGTGGTTCCTATTGCAATCTATGTGTAT	120	
OY	477	tcatgagaagattgaaaacttgaagagctacgcgcgactgtgtatgcccctcagaacat	536	
Db	121	TCATGAGAAGATTGAAAACTTTGAAGGCTACTGCGACGTGTATAGCCCCCTAGAACAT	180	
OY	537	atactgtgtcactgtgagatgagaagctccccaagaactttcaaaagagcggttcaagaacat	596	
Db	181	ATATGCTGTCCATGNTGTGATGAGAAAGTCCCAAGAACTTTCAAAAGAGGCGGTCAAGCAAT	240	
OY	597	tattctgtctcccaaatagtctctcatagccaaagtaagctgtgttcgggtgttatgcttc	656	
Db	241	TATTTCTTCTCTCCCAAAAGTCTTCTTACCACTAGACTGTGTGGGTGGTTATGCTTC	300	
OY	657	ctgttcacaggtgtgcaagctgacctaactgcacatgagaagacttgtccagagctcagtgcc	716	
Db	301	CTGTGTCAGAGGTGCAAGCTGACCTCAACTCAGTAAGAAAGATTCTCCAGAGCTCAGTGGCC	360	
OY	717	gtggaatatctctcctaatacatgtgtggacggacttccctataaagaaacatcaagat	776	
Db	361	GTGGAAATATCTTCTGTGAATACATGTGTGGACGGACTTTCCATATACGACGAATCAGAGAT	420	
OY	777	ggtccaggtctccaagaatgttgaatgggaggaatagcatgagatcagaagttacctcttaa	836	

Db	421	ggtccaggctctc	aaagattgttg	ctgacggagac	ctgacgagctc	ctccta	480
Qy	837	gcacaaagaaccc	gctgtaaatat	ctacttggagtag	tg-agaacacat	-acaccta	894
Db	481	gcacaaagaam	-ccgattggaat	atnactttagag	ttatgcagagac	aatTAACACCTA	539
Qy	895	accacaaagaag	aatgctccccc	cttaataattac	taactgcttac	gaggaatgc	954
Db	540	accacaaagag	cgacgagctc	cccccttatt	aatTAATTAAC	TAATACAGGGAAT	599
Qy	955	attgtagcttc	ccgagattc	gtccacaatg	ttttgaaagacc	cttaatccacaactg	1014
Db	600	attgtagattc	ccccggattat	gcTACAAACAT	ggtttgcaagac	ctTAATACACACAG	659
Qy	1015	attgattggtta	aaagaacact	atatagccagat	gaacacccct	gtggccaccctcag	1074
Db	660	attgactgggt	taaaagaaat	ttatnagccag	atgaacactc	ctgggaacacactt	719
Qy	1075	gcacggtgtag	ctgctgtctg	ttcccaacac	cccccaagta	tcagacatcga	1134
Db	720	caaggtggag	atagccttg	gaacgaac	caaacaccccc	caagaaACATATAC	779
Qy	1135	tctattgcag	gctgctgta	agtgacag	1162		
Db	780	tgactgcac	acgcccac	agacagc	acagc	807	
RESULT 11							
LOCUS	A1587061/c						
DEFINITION	A1587061	671 bp	mRNA	linear	EST 14-MAY-1999		
	tr55902.x1	NCI-CGAP_Pan1	Homo sapiens	CDNA clone	IMAGE:2222258.3'		
		similar to	SW:G6NT_HUMAN	Q02742	BETA-1		
		3-GALACTOSYL-O-GLYCOSYL	-GLYCOPROTEIN	BETA-1.6-N	ACTEYLGLUCOSAMINYLTRANSFERASE	; mRNA sequence.	
ACCESSION	A1587061						
VERSION	A1587061.1	GI:4573502					
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
REFERENCE	1 (bases 1 to 671)						
AUTHORS	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.					
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),						
	Tumor Gene Index						
JOURNAL	Unpublished (1997)						
COMMENT	Contact: Robert Strausberg, ph.D.						
	Email: cgapbs-remail.nih.gov						
	Life Technologies catalog #: 11548-013						
	DNA Sequencing by: Washington University Genome Sequencing Center						
	Clone distribution: NCI-CGAP clone distribution information can be						
	found through the I.M.A.G.E. Consortium/LLNL at:						
	www-bio.llnl.gov/bbrp/image/image.html						
	Insert Length: 2232 Std Error: 0.00						
	Seq primer: -400P from Gldcoo						
	High quality sequence stop: 382						
	POLYA-No.						
FEATURES							
SOURCE	location/Qualifiers						
	1..671						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	/clone="IMAGE:2222258"						
	/clone_id="NCI CGAP_Pan1"						
	/tissue_type="adenocarcinoma"						
	/lab_host="DH10B"						
	/note="Organ: pancreas; Vector: PCMV-SPORT6; Site_1: SalI;						
	/site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.						
	Average insert size 1.72 kb. Life Technologies catalog #:						
	11548-013"						
BASE COUNT	188 a	175 c	131 g	176 t	1	others	
ORIGIN							

Query Match 32.1%; Score 605; DB 9; Length 671;
 Best Local Similarity 97.1%; Pred. No. 6e-165;
 Matches 647; Conservative 0; Mismatches 16; Indels 3; Gaps 3;

```

QY 1168 gagggagacatgataaggtgtctcttattgtctcctgtctctgtctgtgaatccacgagggct 1227
    |||||
DB 671 GAGGAAGACACGATMAAGGTGCTTATGCTCCCTGCTCTGGAAT-CACCAAGGGGGCT 613
    |||||

QY 1228 atctggttatatgaggtctgaggtctgaattgtatgtcttcaaacaccacccgttggcc 1287
    |||||
DB 612 ATGTGGCTTATATGGGCTGGGACTTGAATGATGTC-TCMAAACATCACCTCTTGGCC 554
    |||||

QY 1288 aacaagttgaccccaagtagatgaatgtaatgctcttcgaatgtcttagaagaatacctaagt 1347
    |||||
DB 553 AACCAATTGACCCCAAGGAGATGATGATGCTCTTCAGTGTAGGAAGATACNT-CGT 495
    |||||

QY 1348 tataaagccatctatgtggaatgaattgtgagaacacatagagaagctgtctcctgttg 1407
    |||||
DB 494 TATTAAGCCATCTATGAGGACTGCTTTGAGACACTATGAGAGCGTTCCTACCTGTG 435
    |||||

QY 1408 ggtcaagacatgtacacaacatgtctcagaactgtctgagacagtgtgtgtgagaccag 1467
    |||||
DB 434 GGAAGAAGCATGTACAACATGCTCAGAACTTCTGGAGACTGTGGGTGGAGACCAG 375
    |||||

QY 1468 gcttgcgaatctgtgacatcctttagataagaagggtgctgataatgtgtgtgtaagta 1527
    |||||
DB 374 GCTTGCATATCGTGCAATCCTTTAGATTAAGAGAGCGCTGCTATTAGATGTGGTAAGTA 315
    |||||

QY 1528 gatctttgctgctgcaaatgtcgtcgtggtgtaagtgtgtgttcttcttacccttaacc 1587
    |||||
DB 314 GATCTTTTGTGCTTGCAAAATGTGCTCGGTGTAATCGCTGTCTCTCAACCCCTAACC 255
    |||||

QY 1588 ctatgagttcctcactaacttctcactaagttagaagaataagacgtctgtgataagga 1647
    |||||
DB 254 CTAGTAGTTCCTCCACTAATCTTCTCACTAAGAGATAGAACTGCTGTGATAGGAG 195
    |||||

QY 1648 agtgaaggaaggaatgtgtgtagaagaactgtgaattcaagttgaatgctgtgtgtaagtt 1707
    |||||
DB 194 AGTGAAGGAGGATATGTGTAGAGACATGATTCAGTTGAATGCTGCTGTGATAGCTTT 135
    |||||

QY 1708 tcatctctgtgaggtcgtcgtcttcttaataatctcagaattgtgtgaggtgagagaact 1767
    |||||
DB 134 TCCATTTCTGTGAGCTGCCCTTCTTAATTAATTCAGGTTGTGAGCGTGGAGGAACCTT 75
    |||||

QY 1768 tgaatgaagaagaaccttccctctgtactgttaacttaaaaaataaagctcctgtatc 1827
    |||||
DB 74 TGATGAAGAAGAACTTCCTCTCTGACTGTTAACTTAATAATAATACCTCTGATTC 15
    |||||

QY 1828 aaagta 1833
    |||||
DB 14 AAAGTA 9
    |||||

RESULT 12
BI765410 766 bp mRNA linear EST 25-SEP-2001
LOCUS 603050445F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5190611 5',
DEFINITION mRNA sequence.
ACCESSION BI765410
VERSION BI765410.1 GI:15756988
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 766)
NIH-MGC http://mgi.mgi.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.

```

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM1476 row: 1 column: 12
 High quality sequence stop: 766.
 Location/Qualifiers

FEATURES
 source 1..766
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5190611"
 /clone_11b="NIH_MGC_116"
 /lab_host="DH10B"
 /note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site:1; Nott; Site:2; EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo 2 female, 71 yo male colon; 46 yo female kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH_MGC Library."

BASE COUNT 199 a 184 c 192 g 190 t 1 others
 ORIGIN

Query Match 32.0%; Score 602.6; DB 10; Length 766;
 Best Local Similarity 98.9%; Pred. No. 3.2e-164;
 Matches 638; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

```

QY 1 gatgtctcctcctcacccttccctgtcgtcgtctccacgtctccatctgtgac 60
    |||||
DB 120 GAGTGTCTCTCTCCACCTTCCTGTGCTGCTCCACCTCTCTCCATTTGTGAG 179
    |||||

QY 61 atgttcaatlgagaagactctgacagctgatactgtgtggtctggtctgatact 120
    |||||
DB 180 ATGTTCAATGGAAGAGACTCTCCAGCTGATTAATGTTGGGCTCTGGGCTCTATAG 239
    |||||

QY 121 ctgtctgcaactgtgtctgtgaacttcttcaagttgaagtgtacttgcacacctg 180
    |||||
DB 240 CTGCTGGCAGCTGTGCTCTGAAACTTTCTTCAGGTTGAAGTGTACTGTGACCACTTG 299
    |||||

QY 181 ggtctgagtcacaggaatctcaaaqccagctactgaaagatactctgtataattccgt 240
    |||||
DB 300 GGTCTGAGTCCGAGGAATCTCAAAAGCCAGTACTGTAGAGATATCTTATATTTCC 359
    |||||

QY 241 aaacttcagcaaaagagctatcaactgttcaagggtcaaccgaggagcaagaagca 300
    |||||
DB 360 AAACCTCCAGCAAAAGAGGTCTATCAACTGTTCAGGGGTCAACCGAGGAGCAAAAGAGCA 419
    |||||

QY 301 gttgttcaagctatttctgataaaccctggaggtcaaaagaaggaagccttcaagac 360
    |||||
DB 420 GTGCTTCAGGCTATTTCTGATAACCTGTGAGGTCAAAAGAGGAGAGACCTTTCAACAGAC 479
    |||||

QY 361 acccaactactctcctcaccagagactgtgagacacttcaaggctaaaggaagtccata 420
    |||||
DB 480 ACCCACTACTCTCTCCCTCAACAGAGACTGTGAGCACTTCAAGGCTCAAAAGAAAGTTCA 539
    |||||

QY 421 cagttcccaactgagcaaaagaaggtgaggttccctattgcatctatggtatcat 480
    |||||
DB 540 CAGTTCCTCCACTGAGCAAAAGAGGTGAGTTCCCTATTTGATATCTATGATTCAT 599
    |||||

QY 481 gagaagatggaacttgaagagctactgagagctgtgtatgcccctta-gaactata 539
    |||||
DB 600 GAGCAATTTGAAAACCTTGAAGAGCTACGCGAGTGTATGCCCCCTCAGCAACATATA 659
    |||||

QY 540 ctgt-gttcattgtgagtgaggttcccaagaacttcaag-aggcgttcaagaact 597
    |||||
DB 660 CTGTGTCTCATGTGTGATGAGAGTCCCAAGACACTTTCAAAAGAGCGGTCAAAAGCAATT 719
    |||||

```

Qy 598 attcttgcttcccaaatgcttctcataagccagtaagctgtgtcg 642
 |||||||
 Db 720 ATTCTTCTCTCCCAAAATGCTCTCATACGAGTAAGCTGTGGG 764

RESULT 13
 AM192901/c
 LOCUS
 DEFINITION
 AM192901 603 bp mRNA linear EST 29-NOV-1999
 x155b12.x1 NCI-CGAP Pauli Homo sapiens CDNA clone IMAGE:2678591 3'
 similar to TR:095395 095395 BETA-1
 , 6'-N-ACETYLGLUCOSAMINYLTRANSFERASE. ;, mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL
 COMMENT
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Life Technologies catalog #: 11548-013
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LML at:
 www-bio.lml.gov/bdrip/image/image.html
 Seq primer: -40UP from Glibco
 High quality sequence stop: 372.

FEATURES
 source
 location/Qualifiers
 1..603
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2678591"
 /clone_lid="NCI-CGAP_Pauli"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.72 kb. Life Technologies catalog #:
 11548-013"

BASE COUNT 176 a 158 c 119 g 150 t
 ORIGIN

Query Match 31.8%; Score 599.8; DB 9; Length 603;
 Best Local Similarity 99.7%; Pred. No. 1.9e-163;
 Matches 601; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1225 gctatctgcttctatgagggcgctgggagcttgatgagcttcaaaaccatcacctgtg 1284
 |||||||
 Db 603 GCTATCTCGCTTTATGGGGCTGGGACCTTGATGTGATGCTTCAAAACATCACCCTGTG 544

Qy 1285 gccacaagtttgaccacaagtagatgataatgctcttcagctgtctgaagaataccta 1344
 |||||||
 Db 543 GCCACCAAGTTTGACCCAAAGGTAGATGATGCTCTTCAGTGTAGAGAAATACCTA 484

Qy 1345 cgtataagggcatctatgagggactgaacttgagagacacatgagagcgttgctactctg 1404
 |||||||
 Db 483 CGTTTAAGGCCATTAAGGAGCTAACTTTGAGACACACTATGAGAGCGTTGCTACCTG 424

Qy 1405 tggggcaagagcatgtacaacatgtctagaactgtctggagacagtgctgggtgagagac 1464
 |||||||
 Db 423 TGGGCAAGAGCATGTACAACATGCTCAGAACTGTCTGGGACAGTGGGGGAGAGAC 364

Qy 1465 agggcttgcgaattctgtgcatcctcttaggataagagggctgctatgattgtgtgtaa 1524
 |||||||
 Db 363 AGGCGTTTGCAATTCGTGGCATCCCTTTAGATTAAGAGGGGCTGCTATTGATGTGGGTAA 304

Qy 1525 gtatagcttctgcttgcgaattgctgctgtgggtgaatgctgtctgtctcaccoccta 1584
 |||||||
 Db 303 GTAGATCTTTTGCCCTTGCAAAATGCTGCGGTGATGATGCTGTCTCTCAACCCCTA 244

Qy 1585 accctagtagtcttcccaactacttctcactaagtgtgaatggaactgtctgtgtagg 1644
 |||||||
 Db 243 ACCCTAGTAGTCTCCCACTTAACCTTCTCACTAAGTGAATGAGAACTGCTGTATAGG 184

Qy 1645 gagaatgagagagagatgtgtgtagagcactgttccagttggaatgccgcgtgtgtagc 1704
 |||||||
 Db 183 GAGACTGAAGAGAGGATGTGTGTGAGACACTTGTTCAGTTGATGATCCTGCTGTAGC 124

Qy 1705 ttccatctgtgtgagctgctgtccctcaataatccagttgtgtagcgtgagagaa 1764
 |||||||
 Db 123 TTTTCCATTCTGTGTGAGCTGCGCTTCTTAATTAATTCAGTTGTGAGCGTGAGAGAA 64

Qy 1765 ctltgatgaaagagagacccctccctctctgtactgttacttaaaaaataatagctccga 1824
 |||||||
 Db 63 CTTTGATGGAAGAGAACCTTCCCTTCTGTACTTAACTTAATAATCAATAGCTCCTGA 4

Qy 1825 ttc 1827
 ||||
 Db 3 TTC 1

RESULT 14
 AM662377/c
 LOCUS
 DEFINITION
 AM662377 606 bp mRNA linear EST 06-APR-2000
 h125h05.x1 NCI-CGAP Co14 Homo sapiens CDNA clone IMAGE:2973369 3'
 similar to TR:095395 095395 BETA-1
 , 6'-N-ACETYLGLUCOSAMINYLTRANSFERASE. ;, mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL
 COMMENT
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LML at:
 Image.lml.gov/image/html/lresources.shtml
 Seq primer: -40UP from Glibco
 High quality sequence stop: 380.

FEATURES
 source
 location/Qualifiers
 1..606
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2973369"
 /clone_lid="NCI-CGAP_Co14"
 /tissue_type="moderately-differentiated adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.7 kb. Life Technologies catalog #:
 11531-019"

BASE COUNT 177 a 157 c 121 g 150 t 1 others
 ORIGIN

Query Match 30.9%; Score 582.6; DB 9; Length 606;
 Best Local Similarity 97.5%; Pred. No. 1.9e-158;

Matches	591	Conservative	0	Mismatches	15	Indels	0	Gaps	0
QY	1222	cgggctatcgtcttatagggctgggagcttgaaatgtatgtcaaacaccactg	1281						
Db	606	CGGCTATCCGCTTTATGAGCTGGGAGCTTCATTGCTTCAACACATCACTG	547						
QY	1282	ttggccaacagtttgaaccacaaagttagatgataatgtctctcaatgtcttaagaagaatc	1341						
Db	546	TTGGCTAACCAAGTTTGAACCCATGAGTATGATGAATGCTTCTAGTCTTGAAGACATAC	487						
QY	1342	ctacgtataaaggccatctatgagactgaacttgagacacactatgagagcgcttctac	1401						
Db	486	NTACGCTTATAGGCCATCTATGGAATGACCTTGACACACACTATGAGGCGCTTGCTAC	427						
QY	1402	ctgttgggcaagagcatgtcaacaactgtctcagaactgtctgggacagtggtgggag	1461						
Db	426	CTGTGGGCGCAAGAGCATGTATCAAAACATGCTCAACACTGCTGGGACAGTGGGTGGAG	367						
QY	1462	accagggcttgcattcgtgtgcatcccttaagataagagggctctattagattgtggg	1521						
Db	366	ACCAGGCTTTGCATTCGTCGATCCTTATGATTAAGAGGCGCTGCTATAGATTGTGGG	307						
QY	1522	taagtatatttgccttgcaattgtcgtcgtgggtgaatgtctgtcttctcacc	1581						
Db	306	TAAGTATATCTTTGCTTCGCAATTCGCTGCGGTGAATGCTGCTTGTCTCTCACCC	247						
QY	1582	ctaacctagtagtctccactaacttctcactaagtgaagaatgaaactgtctgac	1641						
Db	246	CTAACCTAGTAGTCTCTCCACATCTTCTCTCACTAAGTGAATGAGAACTCTGAT	187						
QY	1642	agggaagaagaagaagatgtatgttagagacactgtattcgaatgactgtcgtg	1701						
Db	186	AGGGAAGAAGTGAAGAGGATGTGTGAGAGCACTTGATTCAGTTGAATGCTGCTGCT	127						
QY	1702	agcttctcatctgtgtgagctgcgttcccttaataatccaggtttgttagcgtgaaga	1761						
Db	126	AGCTTTTCATCTCTGTGAGAGCTGCGCTCTCTATTAATTCAGGTTGTGTGCGCTGAGGA	67						
QY	1762	gaactttagatgaagaagaaccttcccttctgtactgttaacttaaaaaataagctcc	1821						
Db	66	GAACTTTGATGGAAGAAGAACTTCCTCTCTGTACTGTAACTTAAATAATATAGCTCC	7						
QY	1822	tgattc 1827							
Db	6	TGATTC 1							

FEATURES

High quality sequence stop: 849.

Location/Qualifiers

source

1..864

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4864554"
/clone_lib="NIH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT

235 a 204 c 210 g 215 t

ORIGIN

Query Match

Best Local Similarity

Matches

594; Conservative

29.8%; Score 561.2; DB 10; Length 864;

96.7%; Pred. No. 4.1e-152;

0; Mismatches 18; Indels 2; Gaps 2;

QY

1

gattgttctctccacaccttccctgtgtcgtgtccacactgtctccacattgtgaag 60

Db

252

GATTGTGTCCTCTCCACACTTCCCTGTGCTGCGGTCCACACTGTCCTCCATCTGTGACG 311

QY

61

atgttcaatgga-agaagcttgcagactgcaacttctgtgtgtgtgtgtgtgtat 119

Db

312

ATGTTCAATGGAAGAGAGATCTGCCAGCTGCACTTCTGTGTGTGTGTGTGTGTGTGTGTGT 371

QY

120

gtctgtgcccactgtgtcttgaacttcttctcaggttgaagtgtgactgtgacacct 179

Db

372

GCTGTGCGCACACTGTGCTCTGAACCTTCTTCAAGTTGAAGTGTGACTGTACACACTT 431

QY

180

gggtctggagatccaaggaatcctaagcgaatcgttaggaatatttgtataatttct 239

Db

432

GGGTCTGGAGTCCAGGAACTCAAAAGCCAGTCTGTAGGAATATCTGTATTAATTTCT 491

QY

240

gaacttccaagaagaaggtcttcaactgttcaagggttaccaggaggaaccaagaagc 299

Db

492

GAAACTTCCAGCAAAAGGCTTATCAACTGTTAGGGGTTCACCGAGGGACCAAGAGCC 551

QY

300

agtgctcaggtatcttgaataacctggaaggtcgaagaagaagcgttccacaga 359

Db

552

AGTGCTTCAAGCTATCTGATTAACCTGGAGGTCAAGAAAGCAGAGAGCTTTCACAGA 611

QY

360

caccacactacttccctccacagagactgtgagcaacttcaaggctgaagaagttcat 419

Db

612

CACCCACTACTCTCCCTCACAGAGACTGTGAGCACTTCAAGGCTGAAGGAAGTTCAAT 671

QY

420

acagttcccaactgagaagaagaaggtgtgagttccctattgtatctatcagtgttata 479

Db

672

ACAGTTCCACATGAGCAAGAGAGGTGAGTCCCTATTGCTATCTATATGTTGATTA 731

QY

480

tgaagaattgaaacttgaagaagctactgagactgtgtatgtgtgtgtgtgtgtgtgt 539

Db

732

TGAGAAGATTGAAACTTTGAAAGGCTACTGCAGCTGTGTATGTGCTTCAGAACATATA 791

QY

540

ctgtgtccatgtgtgagagaagtcaccagaacttcaagaagcgggtcaagaactat 599

Db

792

CTGTGTCCCTGTGTGATGAGAGT-CCAGAACTTCCAAAGCAGCGCAACGCAATATAT 850

QY

600

ttctgtcttcccaa 613

Db

851

TTCTGTGTCCAAA 864

Search completed: September 26, 2002, 02:55:07

Job time: 3832 sec

Search completed: September 26, 2002, 02:55:07
 Job time: 3832 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 04:23:02 ; Search time 4629.8 Seconds
(without alignments)
8515.615 Million cell updates/sec

Title: US-09-874-390-1-COPY_436_2319
Perfect score: 1884
Sequence: 1 gattgtgtctctctccacct.....ataataaacttaacaga 1884

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_pat:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_mu:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hgtgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Match Length DB ID Description

1	1880.8	99.8	187275	2	AC092755	AC092755 Homo sapi
2	1833	97.3	2106	9	BC017032	BC017032 Homo sapi
3	1829.8	97.1	2216	6	AX045249	AX045249 Sequence
4	1818.8	96.5	2217	6	AX045251	AX045251 Sequence
5	1818.8	96.5	2217	6	AX045253	AX045253 Sequence
6	1798.8	95.5	2128	6	AF136078	AF136078 Sequence
7	1798.8	95.5	2128	6	AF102542	AF102542 Homo sapi
8	1317	69.9	1317	6	AX087948	AX087948 Sequence
9	1317	69.9	1317	6	AF038650	AF038650 Homo sapi
10	1011	53.7	108873	14	AF318573	AF318573 Bovine he
11	1001.4	53.2	2017	14	AF231105	AF231105 Bovine he
12	960	51.0	4179	14	BHVARBEG	BHVARBEG Bovine he
13	923.8	49.0	1841	10	BC018297	BC018297 Mus muscu
14	726.2	38.5	176483	2	AC096432	AC096432 Rattus no
15	376.8	20.0	2033	10	MMU19265	MMU19265 Mus muscu
16	375.2	19.9	2023	10	D87333	D87333 Mus muscu
17	375.2	19.9	2047	10	D87332	D87332 Mus muscu
18	374.2	19.9	2047	6	AX341786	AX341786 Sequence
19	369.4	19.6	1287	6	AX087946	AX087946 Sequence
20	369.4	19.6	2105	6	I41262	I41262 Sequence 3
21	369.4	19.6	2105	6	I61440	I61440 Sequence 3
22	369.4	19.6	2105	6	I72548	I72548 Sequence 3
23	369.4	19.6	2110	6	AX333255	AX333255 Sequence
24	369.4	19.6	2110	6	AX335500	AX335500 Sequence
25	369.4	19.6	2110	9	HUMGOGG	HUMGOGG Human beta-
26	369.4	19.6	2204	9	HUMGOGG	HUMGOGG Human beta-
27	367.8	19.5	159930	2	AL138757	AL138757 Homo sapi
28	367.8	19.5	183858	9	AL161626	AL161626 Human DNA
29	364.6	19.4	5010	10	S79797	S79797 enzymatic g
30	364.2	19.3	2435	4	BTU41320	BTU41320 Bos tauru
31	307.2	16.3	309	6	AX341869	AX341869 Sequence
32	305.6	16.2	309	6	AX340693	AX340693 Sequence
33	257	13.6	380	14	HSBROVHEZF	HSBROVHEZF M90799 Bovine herp
34	243.2	12.9	3083	9	HUMB16NACT	HUMB16NACT L42230 Homo sapien
35	243.2	12.9	151136	9	HSB189K21	HSB189K21 Human DNA
36	207.8	11.0	189534	2	AC095187	AC095187 Rattus no
37	203	10.8	252504	2	AC073815	AC073815 Mus muscu
38	198.8	10.6	155359	9	AL358777	AL358777 Human DNA
39	198.8	10.6	166788	2	AC021626	AC021626 Homo sapi
40	191.8	10.2	1362	6	AX087935	AX087935 Sequence
41	191.8	10.2	1362	6	AF132035	AF132035 Homo sapi
42	189.6	10.1	3435	9	AB037596	AB037596 Mus muscu
43	188.6	10.0	184590	2	AC093259	AC093259 Homo sapi
44	185	9.8	63853	2	AC087663	AC087663 Homo sapi
45	164.4	9.8	189534	2	AC095187	AC095187 Rattus no

ALIGNMENTS

RESULT 1
AC092755 187275 bp DNA linear HTG 19-JAN-2002
LOCUS Homo sapiens chromosome 15 clone RP11-361D15 map 15q22, ***
DEFINITION AC092755 AC022480
SEQUENCING IN PROGRESS ***
ACCESSION AC092755.3 GI:18249989
VERSION HTG: HTGS_P18249989
KEYWORDS HTGS_P18249989
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 187275)
AUTHORS Rowen, L., Madan, A., Qiu, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dots, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D., and Hood, L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 187275)
AUTHORS Rowen, L., Madan, A., Qiu, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dots, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D., and Hood, L.
TITLE Direct Submission

Db 41378 CTTACGCTTATAGAGAACTACCTACGTTATAGGCCATCTATGGCACTGAACCTTTGAGAC 41437
 Qy 1381 aacatataagagcgttgctacatctgaggcaagacatcatcaaatgctcagaactg 1440
 Db 41438 AACATATAGAGCGTGTGCTACTGTGGGGCAAGACATGTACAAACATGCTCAGACTTG 41497
 Qy 1441 ctggagacagtgctgggtggagagcaaggcttgcacatcgtgacatccttaagataaga 1500
 Db 41498 CTGGAGACAGTGTGGGTGGAGACAGGCGCTTGCATTCGTGGCATCTTTAGATAGA 41557
 Qy 1501 gggctgcatatagattgaggtaagtagatcttggcttgcacatgctcctgggtga 1560
 Db 41558 GGGCTGCTATTAGAGTGTGGGTAGTAGATCTTTGCTCCAAATTCCTCCCTGGGTGA 41617
 Qy 1561 atgctgctgtctctcaccctcaccctagtagtcttctcaccacacttctcactaagt 1620
 Db 41618 ATGCTGCTGTCTCTCTACCCCTAACCTAGTAGTCTCTCCACATACCTTTCTCAGTAGT 41677
 Qy 1621 gagaatgagacactgctgtatagagagagtgaaaggagagatatgtgtagagcaacttgat 1680
 Db 41678 GAGATGAGAACTGCTGTGATAGGAGAGAGATATGAGAGGATATGAGAGCACTTGAT 41737
 Qy 1681 ttcatgtgaatgctgctgctgagtagcttctcacttctgtgagctgctcctcctaataatc 1740
 Db 41738 TTCACTGTAATGCTGCTGTGATGCTTTTCCATTCCTGTGACCTGCTCCATTAATTC 41797
 Qy 1741 caagttgtagagctgagagagcaacttgatgagaaagagacacttctcctcctgtagctt 1800
 Db 41798 CAGGTTTGATGAGCGTGGAGGAGAACTTTGATGAGAAAGAACTTCTCTCTGATCTT 41857
 Qy 1801 aacttaaaataaataagctcctgattcaagatatacctcacttcttgcctagtagatgc 1860
 Db 41858 AACTTAAATAAATAAGCTCTGATTCAGATATACCTTCTTGTGCTAGATATGCC 41917
 Qy 1861 agaaataataaatacctaacaaga 1884
 Db 41918 AGAATATATATATATATATACAGA 41941

RESULT 2
 BC017032 2106 bp mRNA linear PRI 09-NOV-2001
 LOCUS BC017032 Homo sapiens, glucosaminyl (N-acetyl) transferase 3, mucin type,
 DEFINITION BC017032 Clone MGC:9086 IMAGE:3851937, mRNA, complete cds.
 ACCESSION BC017032.1 GI:16877561
 VERSION MGC.
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 2106)
 TITLE Direct Submission
 AUTHORS Submitted (05-NOV-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NIH-MGC Project url: http://mgc.nci.nih.gov
 COMMENT Contact: MGC help desk
 Email: gcgaps-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: http://www-shgc.stanford.edu
 Contact: (Dickson, Mark) mcdpaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Series: IRAC Plate: 20 Row: n Column: 9
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 4758421.
 Location/Qualifiers

FEATURES

source

1..2106
 /organism="Homo sapiens"
 /db_xref="LocusID:9245"
 /db_xref="taxon:9606"
 /clone="MGC:9086 IMAGE:3851937"
 /tissue_type="Colon; adenocarcinoma"
 /clone_id="NIH_MGC_65"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORT6"
 309..1625
 /codon_start=1
 /product="glucosaminyl (N-acetyl) transferase 3, mucin
 type"
 /protein_id="AAH17032.1"
 /db_xref="GI:16877562"
 /translation="MVKRRLCOLHYMALGCVMLATVLAALKSFRLCDSDHGLLES
 RESQSOYCRNLLVNFELPKPARKSINGSGVRGDEAVLOALINLLEVKRREPDTN
 YLSITRCEHFRKARKFIOPPLSKEVEFPVAVSMVHEKTEINERLLRAVYAPONTY
 CVHYDEKSPETFRKAVKAIISCPNVNIVASKLVRYVYASMSRQADLNCHEDLIQSSV
 PWKFLNTGCTDEPKSNAEVOALKMLNKRNSKESVPRKHETRKTHFEVVDTL
 HLNKKRDPPEYNTLMTFGNAYIVASDFYQHLKPKRSOQLLEWVDTSPEHMLA
 TLQRAFMPCSVNHPKRYDIDSMTSIARLKWQHGHEGIDIKGAPAYASGSIHORALCV
 YGAGDLMWMLQNHHLANKRPDKVDNALQCLREYLRKAIYGTSL"

CDS

BASE COUNT 582 a 471 c 516 g 537 t
 ORIGIN

Query Match 97.3%; Score 1833; DB 9; Length 2106;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1833; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gatgtgtcctcctccacatcctcctgctgctgctccacatgctcctcctcctgtagc 60
 Db 249 GATGTGTCTCTCTCCACCTTCCCTGCTGCTCCACCTGCTCCACCTGCTGCTGAGC 308
 Qy 61 atgtttcaatggaagagctctgccaagctgatactatgttggtcctggtcctatg 120
 Db 309 ATGTTCAATGGAAGAGCTCTGCAGCTGATTAATCTTGCGCTGCGCTGATATG 368
 Qy 121 ctgtctgacactgtgctcctgaacttcttctcaagttgagtgagctgacactg 180
 Db 369 CTGCTGGCACTGTGCTCTGAATCTTTCAGGTTGAAGTGTGATCTGCACACTG 428
 Qy 181 ggtctgagctcaggaactctcaagccagctgtaggaatctgtatataattcctg 240
 Db 429 GGTCTGAGTCCAGGAACTCAAAAGCCAGTACTAGAAATATCTGTATAATTTCCG 488
 Qy 241 aaacttcaggaagagctctataactgctcaaggggacccagagggagacaaagagca 300
 Db 489 AAACCTTCAGGAAGAGGTCTATCAACTGTTCAAGGGGTCACCCGAGGACCAAGAGCA 548
 Qy 301 gtgtctcaggtatctcgaataacctgaggtgaggtcagaagaagcgagagccttcacagac 360
 Db 549 GTGCTTCAAGGCTATCTGAATACCTGGAGGTCAAGGAAGGAGAGCCTTTTCAACAG 608
 Qy 361 acccaactacctcctcctcacaagagactgtgagcaactcgaaggtgaaggaagtcat 420
 Db 609 ACCCACTACCTCTCCCTACCAAGAGACTGTGAGCACTTCAAGGCTGAAGAGTAAGTCA 668
 Qy 421 caattccacactgagaagaagagtgagttccctatgatactatgtagtcat 480
 Db 669 CAGTTCCCACTGAGCAAAAGAGGTGAGTTCCTATTGCAATCTATATGCTGATTCAT 728
 Qy 481 gagaagatgaaacttgaagaagctactgtagctgtgtagtgcctccacagaacatatac 540
 Db 729 GAGAAGATTGAAACTTTGAAGAGCTACTGGAGAGTGTATGCCCTCCAGAACATATAC 788
 Qy 541 tgtgtcactgtgtaggaaggtccccaagaacttcaagaaggcggtgaagcaatattc 600

Db 789 TGTGTCATGTGATGAGAGTCCCGAGAACTTTCAAGAGCGGCTCAAGCAATTTATT 848
 QY 601 tctgtctcccaaatgcttcatagcaagtaagctggttcggttggttctatgctccg 660
 Db 849 TCTGTGCTCCCAATGCTTCATAGCCAGTAAGCTGGTGGGGGTTTATGCTCCTCG 908
 QY 661 tcaagggtgaaactgaactccaactgcatggaagacttgctccaagagctgaagtcggtg 720
 Db 909 TCCAGGCTGCAACCTGACCTCAACTGCATGGAAGACTGCTCCAGAGCTCAGTGGCCGTG 968
 QY 721 aaatctctctgatacatgctggaaggaacttctcctaagaagcaatgcaagatgctg 780
 Db 969 AAATACTTCTGTAATCATGTGGAGCGACTTTCCTTAAGAGCAATGCAAGATGCTC 1028
 QY 781 caagctctcaagaatgttgaatggaggaatagcatggaagcaaggtaccctccaagcac 840
 Db 1029 CAGGCTCTCAAGATGTTGAATGGAGGAATAGCATGAGTCAAGAGTACCTCTTAAGCAC 1088
 QY 841 aagaagaaccgctggaataatcaacttgaggtgagagacaatcaactcaaccaac 900
 Db 1089 AAGAAACCCGCTGGAATATCACTTGAAGTGAAGACACATTCACCTAACCAAC 1148
 QY 901 aagaagaagatctccctccctaatttaactatgttccaaggaatgctgctattgtg 960
 Db 1149 AAGAGAGAGATCCTCCCTTATTAATTAATCAATGTTTACAGGGAATGCTACATTGTG 1208
 QY 961 gctccgaagattcgttccaacatgtttgaagaaccctaatacccaacagattgaa 1020
 Db 1209 GCTTCCGAGATTTGCGCCACACATGTTTGAACACCTTAATCCACACACATGATTGA 1268
 QY 1021 tgggttaaagaacattatagccagatgaacaccttggtgcaaccttcagcgtgcag 1080
 Db 1269 TGGGTAAACACATTAATACCCAGATGAACACCTTGGGCCACCTTCAGCGTGCACG 1328
 QY 1081 tggatccctggtccttctcccaaccacccaagctacagacatccagatgacttatt 1140
 Db 1329 TGGATGCTGGCTCTGTGTTCCCAACACCCCAAGTACAGACATTCAGACATGCTTCTATT 1388
 QY 1141 gccaagctggttcaagtggcaggggtcaatgaaggaacatcatalaagggtgctcctatgct 1200
 Db 1389 GCCAGGCTGGTCAAGTGGCAGGGGTCAATGAGGAGACATGATGAAGGTTGCTCTTATGCT 1448
 QY 1201 cccgtgcttggaaatccaacaggggtctatcgttctatgagggctgggagcttgaattg 1260
 Db 1449 CCGTGTCTGGAATCCACGCGGCTATCTGCTTTATGGGGCTGGGAGCTTGAATTTGG 1508
 QY 1261 atgcttcaaaacatcaacctgttggccaagaagtgtgacccaaggatagaatgct 1320
 Db 1509 ATGCTTCAAAACCATCACTGTTGGCCACACAGTTTGACCCAAAGGATGATGATAGCT 1368
 QY 1321 ctccaagtcttagaagaatacctacgttataaggccaatcctatggaactggaacttgagac 1380
 Db 1569 CTTCAAGTGGTTAAGAAATACCTACGTTTAAAGGCCATCTATGGAGCTGAACTTTGAGAC 1628
 QY 1381 acaacttgagaagcgttctgtaacctgttggggaagaagcatgtacaacatgctcgaactg 1440
 Db 1629 ACACATGAGAGAGCTTCTACCTGTTGGGCGCAAGACATGTAACAAACATGCTCAAGACTTG 1688
 QY 1441 ctggagcaaggttgggtgggaggaacagggtcttgcaatttggcactccttaagataaga 1500
 Db 1689 CTGGGACAGTGTGGGTGGGAGCACGAGGGCTTTCAAATCTGCGCATCTTTAGGATTAAGA 1748
 QY 1501 gggctctatagattgttgggtgaagttagatccttgccttgcaaatatgctgctggtga 1560
 Db 1749 GGGCTCTCTATTAATGTTGGGTAAAGTAAATCTTTTGGCTTGGCAAAATGCTGCTGGGTGA 1808
 QY 1561 atgctgcttcttctcaccctcaaccctagtagtctcctccataacttctcaactaagt 1620
 Db 1809 ATGCTCTGTTCTCTACCCCTTAACCTTAGTATGTTCTCTCACTAATTCTCACTAAGT 1868
 QY 1621 gagaatgagaagcgtgtgtataggaagatgaaagggagatatgtgttagagcaactgtat 1680
 Db 1869 GAGAAATGAGAACTGCTGTATAGGAGAGTGAAGGAGGATATGTGTAGAGCACTTGAT 1928

QY 1681 ttcagttgaatgctgctggttagcttccatctctgtgagctgcccgttccctaataatc 1740
 Db 1929 TTCAGTTGAATGCCCTGCTGTAAGCTTTTCATCTGTGGAGCTGCGGTTCTTAATATTC 1988
 QY 1741 caagtttggagcgttggaggaagacttgaatggaagaagaacttcccttcttactgt 1800
 Db 1989 CAGGTTTGGAGCTGTGAGAGAACCTTTGATGGAAGAGAACCTTCCCTTCTACTGTT 2048
 QY 1801 aacttaaaataatagctcctctgattccaagta 1833
 Db 2049 AACTTAATAATATAGCTCTCATTCAAAGTA 2081

RESULT 3
 AX045249
 LOCUS AX045249 2216 bp DNA linear PAT 24-NOV-2000
 DEFINITION Sequence 1 from Patent WO0066727.
 ACCESSION AX045249
 VERSION AX045249.1 GI:11343799
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2216)
 AUTHORS Adolf, G., Heider, K.H. and Sommergruber, W.
 TITLE Tumour-associated antigen
 JOURNAL Patent: WO 0066727-A 1 09-NOV-2000;
 Boehninger Ingelheim International GmbH (DE)
 Location/Qualifiers
 1..2216
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..426
 427..1743
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC17334.1"
 /db_xref="GI:11343800"
 /translation="MWMKRLCOLHYINAGCYMLATVAULKSPRLKCDSPDLGLES
 RESOSQYCNITLYNFKLPARKSINCSVTRDQDAVLAITLNNLEPVKKRPPNDH
 YLSLRDCEHFAERKFTQPLSKSEVEPRIDAVLSWITEKLENNRLLRAYAPNDH
 CVHDEKSPETKEAVKAIISCFPNVFLASKLVRYVASWSVQDLNCEMLDQSSV
 PMKFLNCGDPEPIKTSNAEMVQALKMJNRSMSSEVPKIKETRMKYHFEVVDLT
 HLTNKKDPPPEYNTLMTGNAYIVASRFVQHLNPKSQQLIEWKDYSPDEHLMA
 TLQARLWMPGSPNHPKYDISMTSILARLYKMGHGDIDGAPAPAPCSGIHQRAICV
 YGAGDLNMLQNHHLANKFPKVDNMLQCLIEELRKAIYGTGL"
 1744..2216

FEATURES
 source
 5'UTR
 CDS
 3'UTR
 BASE COUNT 600 a 499 c 547 g 570 t
 ORIGIN

Query Match 97.1%; Score 1829.8; DB 6; Length 2216;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1831; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gattgtgctctctccaccttccctgctgctggttcacacatcctccatctgagc 60
 Db 367 GATTGTGTACTCTCCACCTTCCCTTGGCTGCTGCCACCTGCTCCCATTTGTGAGC 426
 QY 61 atggttcaatggaagagactctgacagctgcatctatctgtggtcttgggtctgataatg 120
 Db 427 ATGTTCAATGGAAGAGACTGCGCAGCGAATTAATTGGGGCTTGCGGCTATATG 486
 QY 121 ctgctggcactgttggctctgaacttcttccaagttgaagtgtgaactctgaacacttg 180
 Db 487 CTGCTGGCCACTGTGGCTCTGAAGACTTCTTCAGGTTGAAGTGTACCTGACCACTTG 546
 QY 181 gttctgagatccaagggaatctcaaacagagtaactgaggaataatctgtataattctgt 240
 Db 547 GGTCTGAGTCCAGGGAATCTCAAGCCAGTACTGTAGGAATATCTTGTATATTTCTGT 606

OY 241 aaactccagcaagaaggtctatcaactgttcaagggtctaccgaggggaccgaagagca 300
 |||||
 Db 607 AAACCTTCAGCAAGAGAGGTCTATCAACTGTTTCAGGGGTCACCCGAGGGACCAAGAGGGA 666
 OY 301 gtgtctcagaggtactctgaataacttgaggtcgaagaagcgagagcccttcacagac 360
 |||||
 Db 667 GGGCTTCAGGGGTATTCGTGAATTAACCTGGAGGTCAAGAGAACGAGAGCTTTCACAGAC 726
 OY 361 acccaactaccctccctcaaccagagactgtgagcacttcaaaaggtcgaagaagttcata 420
 |||||
 Db 727 ACCCACTACTCTCCCTCCACAGAGACTGTGACACTTCAAGGCTGAAAGAGATTCTATA 786
 OY 421 cagttcccaacttgagcaagaaggttgaggtccctattgtcatactctatgtatgtat 480
 |||||
 Db 787 CAGTTCGCCACTGACGAAGAAGAGGTGGAGTCCCTATTGCATACTGATGTGATATAT 846
 OY 481 gagaagatgtgaaacttgaaaggtctacttgagagctgtgtatgtccctcagaacataac 540
 |||||
 Db 847 GAGAAGATTGAAACCTTGAAGGCTACTGCGAGCTGTGTATGCCCTCAGAACATATAC 906
 OY 541 tgtgtccatgtgtgagagagtcctccagaagacttccaagaagcggtccaagaactat 600
 |||||
 Db 907 TGTGTCCATGTGTGATGAGAGAGTCCCAAGAACTTTCAAAGAGCGGTCAACCAATTAT 966
 OY 601 tctgtctcccaaatgtctctcaagctgaagctgtgtcgggtgtgtatgtctctctg 660
 |||||
 Db 967 TCTTGTCTCCCAATGATGTCTTATAGCCAGTATAGCTGTGTTCCGGGTGTTATCCCTCTG 1026
 OY 661 tccaaagtgtcaagctgtgactcaactgtcatgtgaagactgtctcgaagctcaagtcggtg 720
 |||||
 Db 1027 TCCAGAGGTGCAAGCTGACCTCAACTGACGAAAGACTTTCCTCAAGCTCACTGCCCTG 1086
 OY 721 aaatcttccgtgatacatatgtgtgagcgagcttccataaagaagaatcgaagatgtgc 780
 |||||
 Db 1087 AAATACTTCTCTAATACATGTGGAGGAGACTTTCCTATTAAGACCAATCAGATGCTC 1146
 OY 781 caggtctcctaagaatgtgaaatggagaatagcatgagatcagagttactcctaagcac 840
 |||||
 Db 1147 CAGGCTTCACAGATGTGTGAATGGAGAAATACATGAGATCAGAGTACTCTTAAGCAC 1206
 OY 841 aaagaaccccgctgtgaataatactacttgaagtgatgtgagagacaattacaactaacac 900
 |||||
 Db 1207 AAAGAAACCCGCTGGAATAATCATTGTGAGTAGAGACACTTAAACCTTAACCAAC 1266
 OY 901 aagaagaagatcccccctataatctaactatgtttacaggggaatgcgtacattgtg 960
 |||||
 Db 1267 AAGAAGAAGATCTCCCTTATATTTAATTAATGATTACAGGAATCGTACATGTG 1326
 OY 961 gcttcccgagatltcgtccaacatgtlttgaagaaccttaatacccaactgtatgaa 1020
 |||||
 Db 1327 GCTTCCCGAGATTGTGTCACACATGTTTGAAGAACCCCTAATCCCAACACTGATGAA 1386
 OY 1021 tgggttaaaagacattatagcccaagatgaacacctctgtggccaacctttagtgtaag 1080
 |||||
 Db 1387 TGGGTAAAGACATTATAGCCCAATGAACACTCTGGGCAACCTTACGCTGCACG 1446
 OY 1081 tggatgtcgtgtcgtgttcccaacaccccaagtiagacatctcagacatgacttatt 1140
 |||||
 Db 1447 TGGATGCCGCTGCTCTTCCCAACCAACCCCAAGTACGACATCTCAGACATGACTTAT 1506
 OY 1141 gccagagctgtgtcaagtggaaggtcatgagggagacatgataaaggtgtctcctatgt 1200
 |||||
 Db 1507 GCCAGGCTGTGATCAAGTGGGAGGTCATGAGGAGACATGATGAAGGTCTCTTATGT 1566
 OY 1201 cccgtcttggaatcccaacgaggtcatctgtgttattagggcttgggaacttgaaatgg 1260
 |||||
 Db 1567 CCTCTCTGGAATCACACAGGGGCTATCTGCTTATAGGGGCTGGGACTTGAATTGG 1626
 OY 1261 atgtcttcaaaacactgtgtggccaacaagttgacccaagaagtatgatataatgt 1320
 |||||
 Db 1627 ATGCTTCAAAACCATCACTGTTGGCCAACAGTTTGACCCAAAGGTATGATATATGT 1686
 OY 1321 ctccagtgcttagaagaataactacgtataagccaatcatagggactgaacttgagac 1380

Db 1687 CTTACGTGCTAGAGAAATACCTACGTATTAAGGCCATCTAGGACATGAACCTTGAGAC 1746
 |||||
 OY 1381 acacatagagagcgtgtctactctgtgtggcaagagacgtataaacatgtctaagacttg 1440
 |||||
 Db 1747 ACACATATGAGACGTTGCTACCTGTGGGCAAGACATGATACAAACATGCTCAGAACTTG 1806
 OY 1441 ctggagacagttgtgttgagagccagggcttggcaatcgtgtgcatcccttagagtaaga 1500
 |||||
 Db 1807 CTGGGACACTGTGGGTGGAGACAGGCTTTGCCAATTCGTGTGATCCTTTAGATTAAGA 1866
 OY 1501 gggctgtattagattgtgtgtaagtagatccttgccttgcaaatgtcgtcgtgtgta 1560
 |||||
 Db 1867 GGGCTGTATATGATTGTGGTAAGTAGATCTTTGCTTCAAAATGCTCGCTGGGTGA 1926
 OY 1561 atgtcgtctgttctctcaacccttaaccctagtagtctcctcaactcttccactaagt 1620
 |||||
 Db 1927 ATGCTGCTTGTCTCCACCCCTTAACCTTAGTAGTTCCTCCACTACTTTCACATAGT 1986
 OY 1621 gagaatgaaactgtctgtatagagagatggaagagagatatgtgtagaagactgtat 1680
 |||||
 Db 1987 GAGATGAGAACTGCTGTGTATAGGAGACTGAAGAGGATATGTGTAGACACTTGAT 2046
 OY 1681 ttcagltgaaatgcctgtgtgtatgtcttccatctcgtgtgagctgcgttccataaatc 1740
 |||||
 Db 2047 TTCAGTTGAAATGCTGCTGTGTGAGCTTTTCCATTTGTGTGAGCTGCGGTCTTAATATTC 2106
 OY 1741 caggttgtgtatgcgtgtggagagaactttagatgaaagaagaaactccctctgtactgt 1800
 |||||
 Db 2107 CAGGTTTGGTGTACGTGGAGAGAACTTGTGATGAAAGAACCTTCCCTCTGTACTGT 2166
 OY 1801 aacttaaaataatagctcctgattcaagta 1833
 |||||
 Db 2167 AACTTAATAATATATAGCTCCTGATTCAAAGTA 2199

RESULT 4
 AX045251 2217 bp DNA linear PAT 24-NOV-2000
 LOCUS Sequence 3 from Patent WO0066727.
 DEFINITION AX045251
 ACCESSION AX045251
 VERSION AX045251.1 GI:11343801
 KEYWORDS
 SOURCE human.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 2217)
 AUTHORS Adolf, G., Heider, K.H. and Sommergruber, W.
 TITLE Tumour-associated antigen
 JOURNAL Patent: WO 0066727-A 3 09-NOV-2000;
 Boehringer Ingelheim International GmbH (DE)
 FEATURES
 source Location/Qualifiers
 1..2217
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 5'UTR 1..844
 3'UTR 1..426
 CDS 427..999
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC17335.1"
 /db_xref="GI:11343802"
 /translation="MWQKRRCQHLHYMALGCMYLATVAALKISFRLKCDSDHGLGES
 RESQSQCRLNLYNFKLIPAKWSINCSGVTSGDQAVLQALNNLEVKAKKEPTDTH
 YLSLRDCEHFKARKEFLQPLSKSEVEFFPLAYSMVIEHKIENERLLRAYAPONTY
 CVHDEEVPNRFGSGSNYFLPKLHSQ"
 845..1744
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC17336.1"
 /db_xref="GI:11343803"
 /translation="MRRILKTLKGCCLCMLPRTYTVSMNKKSPETFEKAVKAIISCF

KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2217)
AUTHORS Adolif, G., Heider, K.H. and Sommergruber, W.
TITLE Tumour-associated antigen
JOURNAL Patent: WO 006727-A 5 09-NOV-2000;
Boehringer Ingelheim International GmbH (DE)
FEATURES
Source location/Qualifiers
1..2217
/organism="Homo sapiens"
/db_xref="taxon:9606"
5'UTR 1..844
CDS 845..1744
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC17337.1"
/db_xref="GI:11343805"
/translation="MRRLKTLKGYCELCMPRLRYTVSMWKKSPETKEAVKAIISCF
PNVFLASLIVVVVYASMSRVQADLNCMDLLOSVPMKYPLNTGCTDEPKISNAEMV
ALKMLNRNSMSESEVPRKHETPKRKHFEVYRDLHLTNKKDDPPRYLITFTGNAYI
VASRDEVOHLNFKSQQLEWVADTYSPEHLWATLQRAHMMGSPVNHPRYDISDM
TSIARLVKMOGHEGDDIDGAPYAPCSGIHORALICVYGAGDLNMLONHHLANKEDPK
VDDNALQLEIEYLRKAIYGTGL"
1745..2217
3'UTR
BASE COUNT 600 a 498 c 547 g 572 t
ORIGIN
Query Match 96.5%; Score 1818.8; DB 6; Length 2217;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1831; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 gattgtcctccctccacccctcgtgctcggttccacgtgtccctcgttgaag 60
DB 367 GATTGTGACTCCTCCACCTTCCTGTGTGCTGCCACCTGTCCTCCATTCGTGAGC 426
QY 61 atgtgtcaatggaagaagactctgcagactctacttctgtggtcgtgctgctatg 120
DB 427 ATGTTCAATGGAGAGACTGTGCCAGCTGCAATTACTTGTGGCTGTGGCTGCTATATG 486
QY 121 ctgtctgcacactgtgtgctctgaacttcttccagttgaagtgtgacctgacactgtg 180
DB 487 CTGCTGGCCACTGTGGCTCTGAATTTCTTTCAGGTTCAATGTGACTTCACCACTTG 546
QY 181 ggtctgtagtccaggaatctcaaaagcagtaactgttagaataatctgtataattctgt 240
DB 547 GGTCTGAGTCCAGGGAATCTCAAAAGCCAGTACTAGAAATATCTTGTAATTTCTTG 606
QY 241 aaacttcaagcaagaagctctatacaactgttcaaggtgtacccagggaggaacaaagaagca 300
DB 607 AAACCTTCAAGCAAACTGTGCTATCAACTGTTCAGGGGTCACCCGAGGGAGCCAAAGAGCA 666
QY 301 gtgtctcaggtatctctgaataactctgaggtcgaagaagcgaagagccttccacagac 360
DB 667 GTGCTTCAGGCTATTCGTAATACTCGAGTCAAGAGAACGAGAGCCTTTCACAGAC 726
QY 361 acccaactaccttccctcaaccagagactgtgacacttcaaggtcgtgaagaagttcata 420
DB 727 ACCCACTACCTCTCCCTCACCCAGACACTGTGAGCACTTCAAGGCTGAAGAGTTCATA 786
QY 421 cagttccacactgagaagaagaaggtgagttccctatctgactactatagtgtattcat 480
DB 787 CAGTTCCACTGAGCAAAAGAGGTGGAGTTCCTATTGCACTATATGTGATTCAT 846
QY 481 gagaagatgaaactctgaaagctactcgtgagctgtgtatgcccctcaagaacataac 540
DB 847 GAGAAGATTGAAACTTTGAAAGGCTACGCGACGCTGTATGCCCCCTAGAACATATATC 906
QY 541 tgtgtccatgttgatgt-agaagttccccagaaacttcaagaagcggttcaagaacattat 599
|||||

DB 907 TGTGTCCATGTGGATGAAGAACTGCCAGAACTTTCAAAAGAGCGGCTCAAAAGCAATTAT 966
QY 600 tcttgcctcccaaatgtcttcatagccagtaagctgtgtcgggtgtttagtccctcg 659
DB 967 TTCTTGTCTTCCAAATGTCTTATAGCCAGTAACTGTGGTGGGTGTTTATGCTCTCTG 1026
QY 660 gtccaggtgtcagactcgaactcgaactcgaactcgaactcgaactcgaactcgaactcga 719
DB 1027 GTCCAGGGGTGCAAGGTGACCTCACTGATGGAAGCTGTGTCAGAGCTGAGTGGCGTG 1086
QY 720 gaataactcctgtgaataactgttggagcgacttccctataaagaagcaatgcagagatgt 779
DB 1087 GAAATACTTCTCGAATATCATGTGGAGCGACTTCTTAAAGACCAATGCGAGATGTGT 1146
QY 780 ccaggtcctcaagatgttcaatgtggaggaataagcatgtgagtaagagtaaccccaagca 839
DB 1147 CCAGGCTCTCAAGATGTTGAATGGAGGAAATAGCATGAGTACAGTACCTCTTAAGCA 1206
QY 840 caaagaaccccgctggaataatacaactttaggttagtgaagagacaattacacctaaca 899
DB 1207 CAAAGAAACCCCGTGAATATACACTTGGAGTAGAGAGACATATACACTTAACCA 1266
QY 900 caagaagaagaatcctccctcctataatlaactatgttlaagggaaatgtcattatgt 959
DB 1267 CAAAGAAGAGATCCTCCCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1326
QY 960 ggtctcccgagaatctcgtcccaatgtttagaagaacccataatcccaactatgta 1019
DB 1327 GGCTTCCGAGATTTGTCACATATGTTTGAAGAACCTTAATCCCAACACTGATGTA 1386
QY 1020 atgggtlaaaagaacacttataagccagatgaacactctggtccacccctcgaagcgtgacag 1079
DB 1387 ATGGGTAAAGAACACTTATATAGCCAGATGAACACTCTGGCCACCTTCACGCTGACAG 1446
QY 1080 gtgagatgcctgtcctgttcccaaccacccaagaatgaacatctcaagataactctat 1139
DB 1447 GTGAGATGCCTGTGCTGTTCACCAACCCCAAGTACGACATCTGAGATATCACTTAT 1506
QY 1140 tgcagagctgtcgaagtgcaaggtgcatagaaggagaacatcgaatgaaggtgtccctatgc 1199
DB 1507 TCCAGGCTGGTCAAGTGGCAGGGTCTATGAGGAGACATCGATTAAGGTGCTCTTATGC 1566
QY 1200 tccctgtctctggaatcccaagcgggtctatcgtctttaggggcttggggagcttgaatg 1259
DB 1567 TCCCTGCTCTGGAATCCACACGCGGCTATGCGTTATGGGGCTGGGACTGTGAATG 1626
QY 1260 gatgtctcaaaacatcacctcgttggcacaagtttgacccaagaagtatagataatgc 1319
DB 1627 GATGCTTCAAAACCATCACTGTTGGCCMAAAGTTTGACCCAAAGGTGATGATATGC 1686
QY 1320 tcttcaagtgcttaagaagaatacctcgttataagccaactatlggaactgaaacttga 1379
DB 1687 TCTTCAAGTGTAGAGAAATACCTACGTATAGGCACTATAGGCACTGAACCTTTGAGA 1746
QY 1380 cacactatgagacgtctgactaccttlygggcaagacatgtaacaacatgctcaagaact 1439
DB 1747 CACACTATGAGAGCGTGTGTACTGTGGGCAAGAGCATATGCAAAACATGCTCAGACTT 1806
QY 1440 gctggagacagtggtggtggaagacaagggcttgcgaatcgtgtgcaatcccttaagata 1499
DB 1807 GCTGGGACAGTGTGGGTGGGAGACCAAGGCTTGGCAATTCGTGGCATCTTTAGAGATAAG 1866
QY 1500 agggctcattagaatttgggtaagtagatcttgcctcgtcaaatgtcctcgggggtg 1559
DB 1867 AGGGCTGCTATTAATGATGAGGTAAGTAAATTTTGTGCTTCAAAATGTCTCTGGGGTG 1926
QY 1560 aatgtgtctgttctcctcaacccttaaccctagtagtccctcaactcttccactaag 1619
DB 1927 AATGTGCTGTTCCTCAACCCCTTAACCTATAGTCTCCACATCACTTCTCACTAAG 1986
QY 1620 tgaagaatgagaactcgtgtgataaggaagaatgaagaaggagataatgtgtgataagcaactg 1679
DB 1987 TGAGATGGAACCTGCTGTGATAGGAGAGTGAAGAGGAGATATGTGTGATGAGCACTTGA 2046
|||||

Db 1854 ATGCTGCTGTTCTCTCAACCCCTTAACCTAGTACTTCTCCATCACTAATCTTCTCACTAAGT 1913
Qy 1621 gagaatgagaactgctgcatagagagagatgaagagagagatgctgtagagagacttgat 1680
Db 1914 GAGATGAGAACTGCTGCTGATAGGAGAGTGAAGAGGAGATGTGGTAGAGCACTTAT 1973
Qy 1661 ttcaagttgaatgctgctgtagagctttccatctgtgtagctgctgctcctaataatc 1740
Db 1974 TTCAAGTGAATGCTGCTGATAGGAGAGTGAAGAGGAGATGTGGTAGAGCACTTAT 2033
Qy 1741 caggtttgtagctgtagagagagagacttgatgtagaagagagactcctctgtagctgt 1800
Db 2034 CAGGTTTGATGAGAGGAGAGAACTTGATGAGAAAGAGAAC--TTCTCTGTACTGTT 2091
Qy 1801 aacttaaaataaata 1816
Db 2092 AACTTAATAATAATA 2107
RESULT 7
AF102542 2128 bp mRNA linear PRI 31-JAN-1999
LOCUS Homo sapiens beta-1, 6-N-acetylglucosaminyltransferase mRNA
DEFINITION complete cds.
ACCESSION AF102542
VERSION AF102542.1 GI:4204683
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2128)
AUTHORS Yeh,J.C., Ong,E. and Fukuda,M.
TITLE Molecular cloning and expression of a novel beta-1,
6-N-acetylglucosaminyltransferase that forms core 2, core 4, and I
branches
JOURNAL J. Biol. Chem. 274 (5), 3215-3221 (1999)
MEDLINE 99115671
REFERENCE 2 (bases 1 to 2128)
AUTHORS Yeh,J.C., Ong,E. and Fukuda,M.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1998) Glycobiology Program, The Burnham
Institute, 10901 N. Torrey Pines Rd., La Jolla, CA 92037, USA
FEATURES
source location/Qualifiers
1. 2128
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="brain"
/dev_stage="fetus"
354..1670
/note="glycosyltransferase: core 2 Gnt-M"
/codon_start=1
/product="beta-1, 6-N-acetylglucosaminyltransferase"
/protein_id="AADI0824.1"
/db_xref="GI:4204684"
/translation="MVMKRLCOLHYLMALGCMYLLATVALKLSFRLKCDSDHGLLES
RESOYCRNLVNFELKLPARKSINGSGWGDQEAVALINLELKKREPTDTH
YLSLRDCEHFKARFIOPRISKSEVEFPAYVWHEKRENERLRAVYAPONTY
CVHYDEKSPERFKAVKAITISCPNVFIASKLVVYVYASNRVADLNCMDLLOSY
PWKFLNTGCTDFPKSNAEVQALKMLNGNSSESEVPRPKHKTNRKYHEVVDTL
HLTKRKDPPTNYLMTGNAYIVASRDFVOHVKNSQDLIMVAVDTYSDEHLMA
TLORARMPGSPVNHPRKYDISDMSIARLVKMOGHEDDIDGAGAYAPCSGIHQRAICY
YGAGDLMMWLNHLLANKRDPKYDMDNALCLDELYLKKATYGRLEL".
BASE COUNT 569 a 477 c 526 g 556 t
ORIGIN
Query Match 95.5%; Score 1798.8; DB #: Length 2128;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1812; Conservative 0; Mismatches 2; Indels 2; Gaps 1;
Qy 1 gatgtgctctctcaccctctctgctgctgctcaccctgctccatctgtagc 60
|||||

Db 294 GATTGTGCTCCCTCCACCTTCCTCCGTGCTGCCACCTGCTCCCATTTGCTGAGC 353
Qy 61 atgtgtcaatgaaagagctctgccaagctgtaactctgtgagctctgagctgata 120
Db 354 ATGTTCAATGGAAGAGACTCTGCCAGCTGCTTACTTGTGGCTCTGGCTGCTATATG 413
Qy 121 ctgctgccaactgctgctctgaaactcttcttcaagttgaagttgtaactctgacactg 180
Db 414 CTGCTGGCAGCAGTGTGCTGTGAACCTTCTTCAAGTTGAAGTGTGACGTGACACATG 473
Qy 181 ggtctgtagctccaggaatctcaagccaagctgtagaataatctgtataattctctg 240
Db 474 GGTCTGAGGTCCAGGGAATCTCAAGCCAGTACTGTAGAAATATCTTATATTTCTCG 533
Qy 241 aaacttcagaagaagatctataactgctgtaggggtccacccggaggagccaagaagca 300
Db 534 AAACCTCCAGAAAGAGGTCTATCACTGTTCAGGGGTCCCGGAGGAGCCAAAGAGCA 593
Qy 301 gtgtccaggtatctgataaaccctggaagctcaagaagcgaagagccttccagaagc 360
Db 594 GTGCTTCAGGCTATCTGATATTAACCTGAGAGTCAGAAAGAGCGAGACCTTTCACAGC 653
Qy 361 acccaactactctccctcaagagagactgtagaacttcaaggtcgaagagagatcata 420
Db 654 ACCCACTACTCTCCCTCAGAGACTGTGAGCACTTCAAGGCTGAAGAGAACTTCATA 713
Qy 421 cagttccacaggaagaagaggtgtaggttccctattgtcatctatgtagttagtcat 480
Db 714 CAGTTCACATGAGCAAAAGAGGTGAGTCCCTATTTGCTATCTATGATGATATAT 773
Qy 481 gagaagatgaaactctgaaagctagctgtagctgtgtagccctcgaacataatcac 540
Db 774 GAGAGATGTGAATACTTTGAAGAGCTACTGCAGCTGTGTATGCCCCACAGACATATAC 833
Qy 541 tgtgtcatctgtagtagaagatccccaagaacttcaagaaggggtcgaagaactatct 600
Db 834 TGTGTCCATGTGAGTGAAGAGTCCCAAGAACTTCAAGAGGCGCTCAAGCAATATAT 893
Qy 601 tctgtctcccaaatgcttctacagagtaagctgtagtctgtagttagttagttagttagt 660
Db 894 TCTTCTTCCCAATGTCTTCTTCAACAGTAACTGTGCTTCCGCTGTGTTATGCTCTCTG 953
Qy 661 tccaaggtgcaagctgtaactcaactctgtagaagaactctgctcagaagctcaggtcgtg 720
Db 954 TCCAGAGTGAAGAGTGAAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 1013
Qy 721 aaatactctcgaatacatgtagagagagacttctctataagaagcaatgtagagtagtc 780
Db 1014 AAATACTTCTCCAAATACATGTGGAGGACCTTCTTATTAAGAGCAATGCAAGATGCTG 1073
Qy 781 caggtctcagaagatgtagtagagagagatagcagagagtagagtagtacccttaagac 840
Db 1074 CAGGCTTCCAGATTTTGAATGGAGAGATAGCATGGAGTCAAGAGTACTCTCTAAGCAG 1133
Qy 841 aaagaacccgctggaatatalactctgaggtagtagagagacacataaaccttaacac 900
Db 1134 AAAGAACCCTGGAATATACCTTTGAGTAGAGAGACACATTTACACTTAACCAAC 1193
Qy 901 aagaagaagagatccctccctataatctaactagtttaaggggaatgtagactgtg 960
Db 1194 AAGAAAGAGAGATCCCTCCCTTATTAATTAATTAATTAATTAATTAATTAATTAAT 1253
Qy 961 gcttcgcaagtttgcacacatgtagtagaagaccccttaacccaactgtagtaa 1020
Db 1254 GCTTCCGAGATTTGTCACATGTTTGAAGAACCTTAATCCCAACATGATTA 1313
Qy 1021 tggtaagaacactatagaccagatgaacacactctgtagccacactcagcgttagcag 1080
Db 1314 TGGTAAAGACACTTATAGCCCAATGACACCTCTGGCCACCTTCAAGGTGACAGG 1373
Qy 1081 tggatgctgctctgttcccaaccacccaagtagaactctagagatgactatct 1140
Db 1374 TGGATGCTGCTGCTGTCCCAACCAACCAAGTACGACATCTCAGACATGACTTATAT 1433


```
QY 901 aagaataaagatcctcccttaattatactatgtttacagggaatgcgtacattgtg 960
|||||
Db 841 AAGAAAGAGATCCTCCCTTATTAATTAACTATGTTTACAGGAATGCGTACATTGTG 900
QY 961 gctcccgagattcgtccaacatgttttgaagaacccctaatacccacaacgattgaa 1020
|||||
Db 901 GCTTCCCGAGATTTCGTCCACATGTTTTGAAGAACCCTTAATCCACACATGATTGAA 960
QY 1021 tggtaaaagacactatagaccagatgaacacctctgagccacctccctcagcgtgcacg 1080
|||||
Db 961 TGGGTAAGAACAATTATACCCAGATGAACACTCTGGGCCAACCTTACGCGTGACGG 1020
QY 1081 tggatgctgctctgttctcccaaccaccaccaaglaagacatcctcagacatgaacttatt 1140
|||||
Db 1021 TGGATGCTGCTCTGTGTTCCCAACACACCACCAAGTACACATCTCAGACATGACTTCTATT 1080
QY 1141 gccagagctgctcaagtgaggagagtcataaggaggaactgataaggctgctcttaatgct 1200
|||||
Db 1081 GCCAGGCTGGTCAAGTGGCAGGGCTCATGAGGAGACATCATTAAGGGTCTCCTTATGCT 1140
QY 1201 cctctgctctggaatccaccagcgagctatcgtgcttataggctgaggctggaactgattg 1260
|||||
Db 1141 CCTGCTCTGGATTCACACGACGCGGCTATCTGCGTTATGGGCTGGGACCTTGAATTGG 1200
QY 1261 atgcttcaaaaaccatcaacctgttggccaacaagttgacccaaaaggtagaatgaatgct 1320
|||||
Db 1201 ATGCTTCAAAAACCATCACCTGTTGGCCACACAACTTTGACCCAAAGAGATGATTAATGCT 1260
QY 1321 ctccaatgctctagaagaatacctagcttataaggccatcattgaggactggaacttga 1377
|||||
Db 1261 CTTCAGTGCTTAGAAGAAATACCTACGTTATTAAGGCCATCATGAGACTGAACTTTGA 1317

RESULT 10
AF318573 108873 bp DNA linear VRL 18-MAY-2001
LOCUS Bovine herpesvirus 4 long unique region, complete sequence.
DEFINITION AF318573 AF271211
ACCESSION AF318573.1 GI:12802528
VERSION AF318573.1 GI:12802528
KEYWORDS
SOURCE Bovine herpesvirus 4.
ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae.
REFERENCE 1 (bases 1 to 108873)
AUTHORS Zimmermann, W., Brodl, H., Ehlers, B., Buhk, H.-J., Rosenthal, A. and
Golitz, M.
TITLE Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and
JOURNAL Identification of an Origin of DNA Replication
PUBMED J. Virol. 75 (3), 1186-1194 (2001)
1152491
REFERENCE 2 (bases 1 to 108873)
AUTHORS Zimmermann, W., Brodl, H., Ehlers, B., Buhk, H.-J., Rosenthal, A. and
Golitz, M.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2000) Project P24, Robert Koch-Institut, Nordufer
20, Berlin D-10555, Germany
FEATURES
location/Qualifiers
1..108873
/organism="Bovine herpesvirus 4"
/db_xref="taxon:10385"
1..108873
/feature="long unique region; EUR"
complement(445..633)
CDS
/note="ORF B01"
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein pB01"
/protein_id="AAK07921.1"
/db_xref="GI:12802529"
/translation="MEGDYSMAVFGKRTKRLDNLNSAGNEVCALVLRKNINVSAGC
CAVKRNNSFHSLPGKMP"

614..4480
/note="tegument protein; ORF3 BORFA1"
/codon_start=1
/product="v-FGAM-synthase"
/protein_id="AAK07922.1"
/db_xref="GI:12802530"
/translation="MDTSPSIOLCFPVSRAPRLSNCOVIOOFTSGOLPLTSVPLQY
EGVFNVLVGEBOAVREFLVPLDLKLFSSGELTPRESSILKTDSSRLSFYGVPSA
PETALSRLDVALIISGSLAPMSNLTQLIKIVSLIISIKMLNPPSKLHMSLRAL
CPDLSLFRPYKSEIYVGHAFHGAPOGLVSLPDKDLGAEPGLMAGSLMSGSGEMFT
POHGPOTIVNOMCKTAYLVPDVYMTITSTYSLSSGHTDGLAYTGGLDVAARN
LKPLFNSNOMWNEGVNSOLGIASVSTVITSPSSRYELPYTRDIMAIVAMNRC
SLPCHOGVPEPIVESHPLDSVPEGRVAVCSERNTTPTDLPETKRIIVYOLGT
FTSHEDSGREVRCLGALKYLMASFRLPSOVGISAIVKCEETMAKLISVCGTGTGK
ITQSALPEPLNGLLPEFASNRQNMNTMKQHLNIAVIAVIALEAPDNTDTVLK
ACCAIHGCKYKVLGRTRKPHIIVKDKDEQWVRKISLRKSPSPYLSITKRLKOS
ENMODSMEWDTPLSLSTYLLILKHPASEGCKDFIKYKHIDRLSSRAOOGIGARDI
PLSDYILVSDLSIAAAPRGSENMWATRESLDDIPLVNMNEVPGIGALIECPPLS
TYFPIKGAHVAVESLNILVSAFPRKIDVCTPNTITPQTTDSHFGITELMARREP
CSOLGICVFTSCITSNRRGQCLNNSLGIHIVATATACADKVOIGLIDLEPSSI
WMLPISTEHYEPFGTVISQLFRDSPSGHINIDPLVYKLLITVTKHLMNESIISCHDV
CGGLITACFEMAYAGASIALTVPODEDPVFLFISEPGLAVEPVRYVSTVOHLE
TSDIITFDVGRTPSVASNTFTVSKERLIFRESLNEVEMNRHRESTEKOMTYCEY
OCTPEPDPKDIOLHITFPOPSMGPHGPYKHOVNVYLLPCTNPDLSLVALEAGPRVN
LVSTFIDKIVKIVTDTTNVFGICLIGATIEBATLGDRAISMTKHSNVGELKKLI
NSPDVFSLIGHACQILLENKFMGYNKRSNTMTCKENYSLMESRWLNFIPENTH
AVALOSMKGSLBPGWIOGTHLGFAPSESEMYMLSTHGAVATPFGADIASGPALTYP
QNPATGYRTISGLCSADGRHALDLHPGSLNLMOWPHIPKMTPLKVSFWKRMFLDLH
IWANKVREMDOPPEPPLPRLNITKVM"
4608..4811
/note="ORF B02; BORFA2"
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein pB02"
/protein_id="AAK07923.1"
/db_xref="GI:12802531"
/translation="MASPTQYCHRWRYRGPGRKSAKISYGYDRDRTNOLJTSKNAYS
GMSPOCPSCADLNKLFVSOCPLS"
5286..5513
/note="ORF B03"
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein pB03"
/protein_id="AAK07924.1"
/db_xref="GI:12802532"
/translation="MGFNPMVLVYFHCNCTPTSLKHGILIKYIKISYLIRATFLVNV
HICPKPRDPRPGCGPPSLIVGRLAKK"
5943..9347
/note="ORF 6"
/codon_start=1
/product="single-stranded DNA-binding protein MDPB"
/protein_id="AAK07925.1"
/db_xref="GI:12802533"
/translation="MAMKTAQGOQYEDNIGSKAPIGCGYIVYVKKPEPTKEASLIG
NGPFGSVSWSPFLGCTIYERGPFPINVRKAKIDITPTYSVTVYHPRVIMESNDIG
FPPIFHGTLPOLCOETRLDREFGTHSPSSNTINNISDLDGCTNCIDELGLGVIT
EGFKERLYGHMILPLGHVEVOIISPOHAKRILPLDEDLFSKIARDKRRHSDVS
OYLFNSLYTAIAOAIIRKDVAVATVIOAMESQFVRDHKMKRVVKKHDPSSSGHSDG
SLIMDSVASELAVSYGLSFIDAPDMSMLDITWSPLEFVDCITFEDERLOALHANLK
OSIHVNTOLFSTNSMLYLRIOKMPKTKGLDNLVANSYVLOHGLSYSEAOODEYGO
PVPGVOSNLSLGSSTYTHLHLAVASMCNTLARYCYLOPOHOKSONOYSNINJY
VGTAASSDMCNLCQGGKPRVAVCTINTLFLRLDRDFPPIITSHRDPYIITGSVGMYND
ILGNFASFEREKDEGAQVEEVQKTYTWDTLYLRLKLESMGKIDTPPEKLVSDIPSF
IKVFDIDAHVDTEVLKFINCWNNKNNINREKYSIHNVIOYCVWFAQPPCPVQLQ
IYKSLITIIIOVCLPNCMAYEODNSMCGPEMKIMNYQIOLMTNFKGCGPGRVLTG
AELKVMHSDPDCDFPDVDAINGIFSPSTVOYRISRAMLVVKTIKRRIIESNSG
SRAIOSGFLKSSKKDNKTYVGPYAKFLINTYIKIKMFPOTKMSLFMMHFEESKKDPL
ITGVPRDQVLALANTYIENSKLHSELDVLDIIPDLITYAKTKLNNALIRACGQOYF
STTLCLVPEVTVRVEGEYPHVLENEEFGTBOYLKVONRTAQIVQATLKGVDVQM
KRLPIITVPMVINKYTGINGNGIIFHCGKIFGFMGRGDRNLIFENAPFKROSTNAYM
RKHAVEMPIVDNLILKRTVYPSSTEEVNIISRIOSIDACVAPKEDCVSSSLGAVYSE
IGCACADLTADLOLYLGEVYIMSDIELISRIOSIDACVAPKEDCVSSSLGAVYSE
OLEPIGLEQTTTSAFTSEDFQASGLSTIIAAGKKKLNCLMSDFDL"
```


Oy	783	ggctctcaagatgtttgaatgaggagaatagcactggagtcacagagctactcctaagcacaa	842
Db	10804	GGCCCTCAAGATGTTGAAGGGGTAGAACAGTATGGAGTCTGAGGTACCTTCTGATGCCAA	108103
Oy	843	agaaacccgcctggaaataatcaacttggagtgatggagacacattacctaacaacaa	902
Db	108104	AAAAAATGCTGGAAATACCGGTATGAG---GTGACAGACACACTGTACCCCTACACGAA	108160
Oy	903	gaagagagatctctcccttaattactactatgtttcaaggaaatgctaaattgtgac	962
Db	108161	GATGAGGAGACCTCCCTCGTAATATTTACCAATGTTCAAGGGAATGCCCTATTTGTGGC	108220
Oy	963	ttcccgagatcttcgcacaacatgcttttgaagaaccttaatcccaacaactgatgaatg	1022
Db	108221	CTCTCGACCTTTGTGCCAAGATGCTTATGACAAACCTTAATCCAAAGACGTGTTGAATG	108280
Oy	1023	ggtlaaagaacactatagcccagatgaacacctctgggcccaccccttcagcgtgcagtg	1082
Db	108281	GGTCMAAAGACACTTTATACCCCCGAGAACACACTCTGGGGCCACCTTCAGCGCTGCCGG	108340
Oy	1083	gatgcctggctctgtctccaacaccccccaagtagacacttcaagaacatgactctatgc	1142
Db	108341	GATGCTCTGCTCTGTCTCTTACGCCACCCCAAGTATACATTCACATGACTGCCATGCG	108400
Oy	1143	caagcttgcataagtcgacaggtcatalgaggagacatacgaataagggtgtctcctaagtc	1202
Db	108401	CAGGCTCGTCAAGTGGCAGTACACGAGAGGAGATGTCAAGCATGGGGCGCTTATGACAC	108460
Oy	1203	ctgcctctggaatcccaacagcgagctactctgtctatgggctctgggagcttgaaattgat	1262
Db	108461	CTGCTCTGGAATCCATCGGAGGGCCATCTGCAATTTTACGGGGCCGGGACCTGTACTGGAT	108520
Oy	1263	gcttcaaaaccataccctgttggccaacaagtttgaccaaaagttagatgataatgctct	1322
Db	108521	TCTCCAGAACCATCACTCCTTGGCAAAACAGTTTGACCCGAGGAGTGATGATACGTCTCT	108580
Oy	1323	tcaagtgttagaagaataactcaactgataaaggccataatggtactgaaacttggaaac	1382
Db	108581	GCAGGCTTAGAAGAGTACTTACGTACATTAAGGCCATCTATGGGACTGTGAATTTGAGCTGC	108640
Oy	1383	actatgagagcgtctgtacccttggggcaagaagcactgtacaacaactgtccgaacttgc	1442
Db	108641	ACTGT-----GGCCGAGGACTGTGTGCCAATATGCCAGGGGCTACT	108681
Oy	1443	gggacagtgct-gggtgggagaccagggtcttgcattctgtgtccatccttagataaag	1501
Db	108682	GGTACTGTGTGGGGGGGGGAGAGAGGGCTGTGTGAATCATGCAATCCCCACAGGCTCAGAG	108741
Oy	1502	ggctgtatagattgttggtaagtag	1528
Db	108742	GGCTGCTTTGATGGGCTCAGAGATGTTG	108768
RESULT	11		
LOCUS	AF231105	2017 bp DNA linear	VR1_26-MAY-2000
DEFINITION	Bovine herpesvirus 4 beta-1,6-N-acetylglucosaminyltransferase		
ACCESSION	AF231105		
VERSION	AF231105.1	GI:8096688	
KEYWORDS			
SOURCE	Bovine herpesvirus 4.		
ORGANISM	Bovine herpesvirus 4.		
	Vituses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae.		
REFERENCE	1 (bases 1 to 2017)		
AUTHORS	Vanderplasseten,A., Markine-Gorlaynoff,N., Lomonte,P., Suzuki,M., Hirata,N., Yeh,J.C., Bureau,F., Willems,L., Thiry,E., Fukuda,M. and Pastoret,P.P.		
TITLE	A multipotential beta-1,6-N-acetylglucosaminyl-transferase is encoded by bovine herpesvirus type 4		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (11),	5756-5761 (2000)

[illegible]

QY 483 gaagattgaaacttgaaaggctactgagactgtgatagcctccagaacatactg 542
 DB 658 GAAGATTGAGAACTTGAACGGCTCTCGACCCCTGTATGCCCTCAGACATATCTG 717
 QY 543 tgcacatgtgataggaagctcccgaaacttcaaaagagcgctcaagaacattatc 602
 DB 718 TGTCCACGTGATGATGAGTCCCGCAGAGACTTTCAAGAGCGGTCAAGGCATATTTC 777
 QY 603 ttgcttcccaaatgtcttcataagcagtaagctgttgggtgtgtttatgctctgtc 662
 DB 778 CTGCTTCCCAATGTCTTATGCGCAGTAAGTTGGTCCGGGTGTATGCTCTCTGTC 837
 QY 663 caaggtgcaagctgaccccaactgcatggaagactgtccagaactgacgtcggtgaa 722
 DB 838 CAGAGTGAAGCTGACCTGAAGCTGTATGGAACCTTGCTCCAGACTCATGTCTGGA 897
 QY 723 ataactctgaatacatgtggagcgacttctcctaagaagcaatgcagagatgtcca 782
 DB 898 GTACTTACTGAATACATGCGGACAGACTTCCCATTAAGAACCAATGCCGAGATGTCT 957
 QY 763 gactcgaagatgtggaatggaggaatagcatgagatcagaaggtactcctcaacaaa 842
 DB 958 GGCCCTCAGATGTTGAAGGGTAAAGACGATGAGAGTCTGAGTACCTTGTAGTCCAA 1017
 QY 843 agaaaccgctggaataatacaacttgaagtagtgagagacacatacaactcaacaaca 902
 DB 1018 AAAAAATGCGTGAATACCGCTATGAG---GTGACAGACACACTGTACCTTACCAAGAA 1074
 QY 903 gaagaagatcctcccttataatctaactatgttcaaggaatgcgtacatgtgac 962
 DB 1075 GATAAAAACCTCCCTCGATTAATTTTACCCATGTTTCAAGAGGATGCTATTGTTGTC 1134
 QY 963 ttccgagatttcgcaccaatggtttgaagaaccttaatcccaactatgtaatg 1022
 DB 1135 CTCTGAGCCTTGTTCACACATGTCTTGAACACCTTAATTCCTCAATATGCTGGAATG 1194
 QY 1023 gataaagaacacttaagccagatgaacaacctcgggcccactcagcggtgacggtg 1082
 DB 1195 GGTCAAAACACATATACCGCCGAGAACACCTCTGGGCGACCTTCACGCTGCTCCG 1254
 QY 1083 gatgcctgctctgttcccaaccaccccaagtagacatctcagaatgacttataatg 1142
 DB 1255 GATGCGCTGCTGTCTTCTTACGACCCCAAGATATCATCTCAGACATGATGCCATGCG 1314
 QY 1143 caagctgttcaagtgacgggtgtcattgagaggaacatcgataaggtgtcctctatgctcc 1202
 DB 1315 CAGGCTCTCAAGTGGCAGTACACAGAGGAGATGTCAAGCATGAGGGGCCCTTATGCAAC 1374
 QY 1203 ctgccttgaaatcccaagcggtcatctgctttatggagctggggaactgtaattgat 1262
 DB 1375 CTGCTCTGGAATTCATCGGAGGGCCATCTGCAATTTACGGGGCCGGGACCTTACTGGAT 1434
 QY 1263 gcttcaaaacacatccctgttggccaacaagtttgaccacaaggtatagatgactct 1322
 DB 1435 TCTCCAGAACCATTCACCTTTTGGCAACAAGTTTGACCCGAGAGGTGATGATACGTCT 1494
 QY 1323 tcaagctttagaagaatacctacgttataagggccatctatggagactgtaacttggaaac 1382
 DB 1495 GAGGTGCTTAAAGAGTACTACATACGATTAAGGCGCATTTATGAGTGAACCTTTGA----- 1549
 QY 1383 actaagagagcgttctacacgttgggcaagagcatgttacaacaatgctcagaacttct 1442
 DB 1550 -----GCTGCACGGTGGGCCAGGGACTGTGTCACAACTGCCACAGGGCGTACT 1595
 QY 1443 gggacagtgct-gggtggagagcaagagcttggcaatctggtggaccccttaagataagag 1501
 DB 1596 GGTACTGTTGGGTGGGGAGAGAGGCTGTGGAATTCATGCAATGCCCAAGGTGACAGAG 1655
 QY 1502 ggtgctctatgatgtgtggtgaagtag 1528
 DB 1656 GGTCTCTTCGATGGGGGTGACGATGTTG 1682

RESULT 12
 BHVAREGF
 LOCUS BHV4REFC 4179 bp DNA linear VRL 06-JUL-1995
 DEFINITION Bovine herpesvirus type 4 DNA for nonconserved region F (DN599 like strain).
 ACCESSION 246386
 VERSION 246386.1 GI:562228
 KEYWORDS
 SOURCE
 ORGANISM
 Bovine herpesvirus 4.
 Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae.
 Lomonte, P.
 1 (bases 1 to 4179)
 DIRECT SUBMISSION
 JOURNAL Submitted (20-OCT-1994) Patrick P.L. Lomonte, Virology, University of Liege, Faculty of Veterinary Medicine, Bd de Colonster, Sart Tilman, Liege, 4000, BELGIUM
 REFERENCE
 AUTHORS Lomonte, P., Bublolt, M., van Santen, V., Kell, G.M., Pastoret, P.P. and Thiry, E.
 TITLE Analysis of bovine herpesvirus 4 genomic regions located outside the conserved gammaherpesvirus gene blocks
 JOURNAL J. Gen. Virol. 76 (Pt 7), 1835-1841 (1995)
 MEDLINE 97201604
 FEATURES
 source location/Qualifiers
 1. 4179
 /organism="Bovine herpesvirus 4"
 /db_xref="taxon:10385"
 /clone="EcoRI H (Bublolt et al., 1990) tot"
 BASE COUNT 1192 a 927 c 951 g 1109 t
 ORIGIN
 Query Match 51.0%; Score 960; DB 14; Length 4179;
 Best Local Similarity 80.9%; Pred. No. 3.3e-295;
 Matches 1180; Conservative 0; Mismatches 245; Indels 33; Gaps 4;
 QY 72 gaagagactctgcacgcagcatctacttctgtggtccttggtgctctatagctctgccaac 131
 DB 2649 GAAGAAGCTTTGGCCGGGACACATCTGTGGGCCCTGTATGCTGTGGCCGT 2708
 QY 132 tgtgctctgaacttcttctcagggttgagtgacttgacacactgggtcgtgagtc 191
 DB 2709 TGTTCCTGAGACTTTTCCCTCAGGTTTAATGTATGATGATGATCTTGGATCTGAGTGC 2768
 QY 192 caaggaaatcctcaagccagactgtaggaatactctgtataaatttccctgaactcagc 251
 DB 2769 CAGGAGCTTCCAGAGTCAAGCATGTAGGACATGTTGTATCAATATGCTGAACTCCAGC 2828
 QY 252 aaagaggtctatcaactcttcaaggtgtcaccggaagggaaccaagggcaggtctcagc 311
 DB 2839 AAAGAGATCCATCAACTCTTCTGTGGGATCACGAGAGGACACGAGAGCATGTGTCAGGC 2888
 QY 312 tatctgataaacctcgaggaggtcagaagaagagagagccttcaagaacaccactact 371
 DB 2889 CCTCTAGACAACTTGGAGATGAAGAAAAAGCGGTGCGCTCTCACTGGCACCTATTACTT 2948
 QY 372 ctccctcacaagagactgtgagacactcaaggtctgaaggaagttcatatagttccact 431
 DB 2949 CAACATTAACCAAGAGACTGTGAGCGCTTCAAGGCCCAAGGAAGTTCATATACGTTCCACT 3008
 QY 432 gagcaagaagaagtgagagcttccctatgcatctctatggtatgcatatgagaagatga 491
 DB 3009 GAGTAAAGAGAGCTAGACTTCCCATTTGCTTACCTGATGTGGTGCATAGGAATATGA 3068
 QY 492 aaacttgaagagctacgcagagctgtgtatgcccctcaagaacatactgtgtccagt 551
 DB 3069 AAACCTTGAACGGCTGCTGCGAGCGGTGTATGCCCTTCAAGAACATATATCTGTGTCAGCT 3128
 QY 552 ggaatgaagatcccaagaacttcaagaagcggttcaagaagcaatattcttctgctcc 611

Db 3129 GGAATGCAAGTCCCGACAGACTTTCACAAAGGCGGTCGAAGCCATTATTTTCCTGCTTCC 3188
 QY 612 aaatgctctatagccagtaagctggtcaggtgttatagtctcctcgtgcagagtgca 671
 Db 3189 CAATGCTTCAATGCGGAGTAAGTGGTCCCGGAGGTGTTATGCTCCGTCGACAGAGTGA 3248
 QY 672 agctgacctcaatgctatgtaagaagactgtctcagagctcagtgccgtggaatacttct 731
 Db 3249 AG-----CTGATGGAAGACTGTGCTCCAGAGCTCAGTGCACAGGAAGTACTACT 3298
 QY 732 gaatacattggtgagcagacttctctataagagacatgacagatgtgtccagctctca 791
 Db 3299 GAATACATGCGGAGACAGACTTCCCATTAAGACCAATGCGAGATGCTCTGCGCTTCA 3358
 QY 792 gatgttaatggtgaggaataatgacatgagatcagatcctcctaaagacaagaacccg 851
 Db 3359 GATGTGAAGGTAAGAAAGTATGAGTGTGAGTACCTTCTGAGTCCAAATAAATAATCG 3418
 QY 852 ctggaataatcactttgaggtagtgagagacatcaccctaaacaaagaagaaga 911
 Db 3419 CTGGAATATACCGTATGAG--GTGACGACACACTGTACCTACAGCAAGATGAAGA 3475
 QY 912 tccctccctcaatctaattactatgtttacaggaatggtacatgtgtgtctccgaga 971
 Db 3476 CCCTCCCTGATATATTTACCATGTTTACAGGGAATGCTTATTTGTGCTCTGAGC 3535
 QY 972 ttctgtcccaacatggtttgaagaacccctaaatcccaacatgattgaatggttaaaga 1031
 Db 3536 CTGTTGCAACATGCTTTATGACACCTTAATATCCCAAGACTGTTGATGGGTCAAGA 3595
 QY 1032 caattatagccagatgaagaactctctgagcaccctcagcgtgacagtgagtgatctg 1091
 Db 3596 CACCTTATAGCCCGGAGAAACCTGTGGGCCCTTCACAGTGCCTCGTGATGCTGG 3655
 QY 1092 ctctgttcccaacaccccaagtaacagacatcagacatctctatctcagagctggt 1151
 Db 3656 CTGTGTTCCATAGCCACCCCAAGATATCATCTCAGACATACCTCCATGCCAGGCTG 3715
 QY 1152 caagtggcaggtgacatgagggagacatgaaaggtgtcctctatgtctcctgctcg 1211
 Db 3716 CAAGTGGCACTTACACAGAGGAGATGACATGAGGGGGCCCTTATGACACCTGCTGG 3775
 QY 1212 aatccacacaggggactatgagctttatgaggtctgaggtgacttgatgagtgctaa 1271
 Db 3776 AATTCATCGAGGAGCCATGCTGCAATTTACGGGGCCGAGGACCTGTACTGAGTTCCAGAA 3835
 QY 1272 ccaatccctgttggccaacagttgacccaagaagtatgataatgctctcagtgct 1331
 Db 3836 CCATACACCTCTTGGCAAAAGTTTACCCGAGGGTGTGATGATTAACCTCTGCAGTCTT 3895
 QY 1332 agagaataactcagtataagccatctatggaactgaaacttgagacacactatgaga 1391
 Db 3896 AGAAGAGTACTGCTGCTATGAGGCCATATGAGGACCTTATGAGCTTCACTGCTG--- 3951
 QY 1392 ggcgttactactgtgggagcagagcaltgtaacaacatgctcagaactgtgtggagac 1451
 Db 3952 -----GGGCCAGGAGACTGTGCCAACAATGCCCCAGGCGCTACTGTGACTGTG 3996
 QY 1452 t-gggtggagacagaggtcttgcacatctggtgacatccttataagataagagggctg 1510
 Db 3997 TGGGGTGGGGGAGAGGCGCTGTGAATTCATGTCATCCCGAGGTCAGAGGCGCTGTT 4056
 QY 1511 tagattgtgggtaagtag 1528
 Db 4057 GATGGGTCTAGATGTTG 4074

RESULT 13
 LOCUS BC018297 1841 bp mRNA linear ROD 06-DEC-2001
 DEFINITION Mus musculus, RIKEN cDNA 2010013H22 gene, clone MGC:18826
 IMAGE:4208500, mRNA, complete cds.
 ACCESSION BC018297

VERSION BC018297.1 GI:17390702
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1841)
 REFERENCE
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (03-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: villalob@bcm.tmc.edu
 Villalob, D.K., Luna, R.A., Hale, S.M., Huliy, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 24 Row: p Column: 13
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.
 Location/Qualifiers
 1..1841
 /organism="Mus musculus"
 /db_xref="locusid:7207"
 /db_xref="taxon:10090"
 /clone="MGC:18826 IMAGE:4208500"
 /tissue="Colon, normal, 5 month old male mouse."
 /clone.lib="NCI CGAP Co24"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORT6"
 192..1388
 /codon_start=1
 /product="RIKEN cDNA 2010013H22 gene"
 /protein_id="AAH18297.1"
 /db_xref="GI:17390703"
 /translation="MDLSEFEQSOYCRDLTKTLKLPKSSINSGVIRGOKAYTO ALLNLEIKKKOOLTEADYLRMTADCEHFKRKFIOVPLSKESAPFIAISVYHE KLENEERLRAVYTPQNYCVHMDKSSPEFQAVRAIVSCPFPNVLGSKLVSYAS WSRVADLNCMEDLQSPVPMKYLTNCTGTPDKITNEMVVALKLLGONSMESEVP PHKSKRMKYHYEVDTLHMTSTKRPPTNNLTFTGNAYVASDFLEHYNSKAR OLIEWVKDYSPDEHLMATLORASMMPSDDPLHRREFDLSDMAIRIRTKWYHEDIE NGAPYTSKSGIHQRAVCYVSGSDLHWIILNHLANKRPKVDNVLOCLEYLNHKA ILYTEL"
 BASE COUNT 496 a 448 c 469 g 428 t
 ORIGIN
 Query Match 49.0%; Score 923.8; DB 10; Length 1841;
 Best Local Similarity 73.1%; Ped. No. 1.e-283;
 Matches 1307; Conservative 0; Mismatches 447; Indels 35; Gaps 8;
 QY 46 tcccatctgtgagcagtggttcaatggaagacatctgcagctgacttactgtggt 105
 Db 60 TCTCTGCTGTGAAGATGACTTCTGCGACAGGCTCTGCTGCTATGATGCTGTGACC 119
 QY 106 ctgagctgctatagctgctgagccactggtgagcctgaactcttcaagtgtagt 165
 Db 120 CTGGTGTCTACTAGCTACTGCGCAATCTTCCCTGAATACTGTCCCTCAGACTGAAGTGT 179

OY	166	gacccctgaccacttgggtctgagctccagggaactcaaggccagctacgtlaaggaatac	225
Db	180	gactttccatgcacatgcatgtgtcgacactctggaggaaatttcaaaagccactactgcagaggaattctc	239
OY	226	ttgatataattccctgaaacttccagaaagaggtcatcaactcttcaagggtgaaccgca	285
Db	240	ctgtacacaaagaccttgaagactgcccagccaaagatttccatcaactgcgtcaggggtcatttga	299
OY	286	gggggacaagaagggagctgctcagagctatcttgaataaccttgaggtcaagaagaagcga	345
Db	300	ggggagacgaagaagcggtgaccacagggctgctgaataaacttgaataattgaagaagacag	359
OY	346	gagccttccacagagaccaccaactacctctccctccacagaagactgtgagcaactcaagct	405
Db	360	cagctcttccacagagcccgacttactttagagatgacagcagactgtagcacacttcaagacc	419
OY	406	gaaaggagaagtctacaggttcccccagagcaagaagaaggtgaggttccctattgataac	465
Db	420	aagaggagaagttttatnacgggtccacactgacacaaagaaagagccaggttcccattgcccgtac	479
OY	466	tcatagttgattca tggagaagattgaaaacttgaagaagctacgtcgagcgtgtgatacc	525
Db	480	ttccatgtgtgggtgacttgagaagattggaamaacttgcgaaggttctctcgagctgtgtacacc	539
OY	526	cctcagaacacatactgtgtccatgtgtga tggaaagtlcccccagaacttccaagaagcg	585
Db	540	ccttcagaaatgtaftacgtgtgacacatgaaatgaaatgttgcagaaaccttttaagcagcca	599
OY	586	gtcaagaacatlatcttcttcttcccaaatgtctctataagccaaatgaactgtgttcggtg	645
Db	600	gtcagggcccatcgtgcatactcttcccacattgtcttcatgctatgaaagttggctgctacgtg	659
OY	646	gtttatgctcccttggtcccaagggtgagaaagcttgcacccaactgata tggaaagacttgcacag	705
Db	660	gtctatgcttcttcgtgtcctcaggggtcagcgtgacacttaactgcgatggaagacttgccttcag	719
OY	706	agctcagtgccgttggaaataactcttcctgaaatacgtgtggagcaggaacttccctataaagac	765
Db	720	agccccgcggccatgaaatgtaactctctgaaacactcgtgggacagacttttccatcaaaacc	779
OY	766	aatgcagagatgtgtccaggtcttccaaagtgttgaat tgggaaggaataagcatgtgagtcagag	825
Db	780	aatgctgagagtggtcaaggcccttcaagcgtatggaagggcacaamaagatggagtcagacag	839
OY	826	gtactcctcctaaagcaagaagaccgctggaatactacacttggaggttagtgaagaagca	885
Db	840	gtaccaccacttccacacttataaaaaatccccgtgaaatattacattatgac---gtgcacacacaca	896
OY	886	ttacaccttaaccaagaagaagagatctctcccccttataatttactatagtlttcaaggg	945
Db	897	tttgacatgacacacagaaagaaagacggccgccacttaataatcctaattggttcacgtgg	956
OY	946	aatgcgtacatgtgtgcttcccgagatlttcgtccaaacatgtlttgaagaaccctaaatcc	1005
Db	957	aatgcctaacagtggtgctgttctcgacagcttctattgaacacagtttgaatcaaaamaacc	1016
OY	1006	caacacagatgtgaatgggtcaaaagcaactatagccagatgaaacacactcgtgggcaacc	1065
Db	1017	cggcacactgattcgaatgggtggttaaaagacacactatagctccgattgagcacacttggggccacc	1076
OY	1066	ctcaagcgttgaacgltgagatcctgtgctgttcccaacaccccacaatgtacagacatctca	1125
Db	1077	cttcacagcggtgcctgtggagatgcttggaatcccttgcattcgaaattttgacctgtca	1136
OY	1126	gacatgacttctatctccaggtctgtcgaagtgtgcaggtgtcatatgagggagacatgataag	1185
Db	1137	gacatgaaagacatttgggaacacttaacacaaatgtgtacagacattgagagagacatttgagAAC	1196
OY	1186	ggtctctctctgtctccctctctctgaaatcccaacagcggtcatctctgcttataatgggt	1245
Db	1197	ggggcacccttaccgtcttctgtcctaggaatnccacacacgggctgtctgtgtttatggcgca	1256
OY	1246	ggggacttgaatttgaatgcttcaaaacctacactgttggccaacaaglttgaaccacag	1305

Db	1237	GGGAGCCCTGCACCTGGATACCTTCAGAACCAACACACCTTTGGCCAAAGATTGACCCAAAG	1316
Oy	1306	gtatgatataatgtctctcttgatgtctcttagaagaatctactgctctataagaagcaatctatgg	1365
Db	1317	GTGATGATATATGTTCTTCAGTGTTTAGAAAGATATTTACGTACAAAGCCATCTATATGG	1376
Oy	1366	actgaactltgagacacatctagagcyltctactcyltggtgccaagacatctatcaaa	1425
Db	1377	ACTGAACCTATGAG--ACACTGAAGACTATGTTTACCTCTGTGGCAAGAAATG-----	1428
Oy	1426	catgtcagaactctgtctgcagacatgtlgtgttggaagcaagggcttccaaltctgtgca	1485
Db	1429	-----CCAGAGCCTACTGGGGGAGTGTGGGGGTAGATAGCTGGGGGCTAGACOTCCATGTG	1483
Oy	1486	tccctttagabaaagagggctgtctataagatltgtgttaagtatcctlttgcttgcaa	1545
Db	1484	TCCCTTAGAGATGAGGGGGCTGTATTGTGACAGGGGGTAAGTAGACCTGTGCTCTTGGA	1543
Oy	1546	tttgtgcctgtgggtgaatgctgtccttgttctctcaaccccttaacctagttcctcacta	1605
Db	1544	-GCTGTGCGCTGGGAAGGGGTGCTATCTTCACAGCTCCCTCCCAA-----CTTCTTGG	1594
Oy	1606	actttctcactaaagtgaatltgaactcgtctgtgaatgg-----agatgtaagsgagat	1661
Db	1595	GCTTCTCTGCGAGACTCTGAACATGAATATCTTTACTTGGGATATTAAGGGGGGACAT	1654
Oy	1662	atgtgttagaggaacttgaattcaattgaatgcctctgttagcttttcatctctgtgag	1721
Db	1655	GTGACACAGGACCTTGATTTTGGCTGAAAGACTTCTGTCAAGTT-----CTCTACAGAG	1709
Oy	1722	ctgcgcgttccctaattccagatcttvgtaagctgtggaagagaaactltgatagtgaagaa	1781
Db	1710	ACTTTGTTTCAAAACTCTGAGGGGGTGTGCTCTGGGTGGGAACCTTAGTSCAAAGAGC	1769
Oy	1782	ccttcctcctctgtactgttactactaaataaataagctccctgattcaaa	1830
Db	1770	CCCCCTCTTATATGTGTAA--TTTAACTAATAAGACTCTTAACCAAA	1817

	RESULT	14
	LOCUS	AC096432/c
	DEFINITION	Rattus norvegicus clone CH230-36K24, *** SEQUENCING IN PROGRESS
	ACCESSION	AC096432
	VERSION	AC096432.4 GI:17946738
	KEYWORDS	HTG; HTGS_PHASEI.
	SOURCE	Norway rat.
	ORGANISM	Rattus norvegicus Eumalvota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 176483)
	REFERENCE	Murphy D.M., Adams C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbarta,J., Benton,J., Blimage,K., Blankenburg,K., Bonnam,D., Bouck,V., Burke,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burnch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Catterl,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdry,I., Christopoulos,C., Cleaveland,C.D., Cox,C., Coyle,M.D., Datborne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Deedrich,D.A., Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite K.J., Draper H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Fliegls,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Geo,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunatratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J.S., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulysk,S., Hunee,d., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Joinet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, W., Louised, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, M., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rotjokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Tellrod, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 176483)
Worley, K. C.

Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17062249.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GFPK
Center clone name: CH230-36K24
----- Summary Statistics
findthraplast
Assembly program: Phrap; version 0.990329first call to
findthraplast
Consensus quality: 145213 bases at least Q40
Consensus quality: 154367 bases at least Q30
Consensus quality: 161818 bases at least Q20
Estimated insert size: 148877; sum-of-contrigs estimation
Quality coverage: 0x in Q20 bases; agarose-ip estimation
Quality coverage: 2.1x in Q20 bases; sum-of-contrigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 64 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
10789: contig of 10789 bp in length
10790 10889: gap of unknown length
10890 17590: contig of 6701 bp in length
17591 17690: gap of unknown length
17691 23617: contig of 5927 bp in length
23618 23717: gap of unknown length
23718 27572: contig of 3855 bp in length
27573 27672: gap of unknown length
27673 31919: contig of 4247 bp in length
31920 32019: gap of unknown length
32020 37411: contig of 5392 bp in length
37412 37511: gap of unknown length
37512 41536: contig of 4025 bp in length
41537 46655: gap of unknown length
46656 46755: gap of unknown length

46756 51783: contig of 5028 bp in length
51784 51883: gap of unknown length
51884 55562: contig of 3679 bp in length
55563 55662: gap of unknown length
55663 59942: contig of 4280 bp in length
59943 60042: gap of unknown length
60043 63342: gap of unknown length
63343 63442: gap of unknown length
63443 69123: contig of 5581 bp in length
69124 73096: contig of 3973 bp in length
73097 73196: gap of unknown length
73197 77220: contig of 4024 bp in length
77221 77320: gap of unknown length
77321 80222: contig of 2902 bp in length
80223 80332: gap of unknown length
80333 84643: contig of 4321 bp in length
84644 84743: gap of unknown length
84744 87894: contig of 3151 bp in length
87895 87994: gap of unknown length
87995 89819: contig of 1825 bp in length
89820 89919: gap of unknown length
89920 93241: contig of 3322 bp in length
93242 93341: gap of unknown length
93342 95561: contig of 2220 bp in length
95562 95661: gap of unknown length
95662 98293: contig of 2632 bp in length
98294 98394: gap of unknown length
98394 99749: contig of 1256 bp in length
99749 102029: gap of unknown length
99750 102029: contig of 2280 bp in length
102030 102129: gap of unknown length
102130 105439: contig of 3310 bp in length
105440 105539: gap of unknown length
105540 107561: contig of 2022 bp in length
107562 107661: gap of unknown length
107661 110897: contig of 3236 bp in length
110898 110997: gap of unknown length
110998 114069: contig of 3072 bp in length
114069 114170: gap of unknown length
114170 116366: contig of 2197 bp in length
116366 116465: gap of unknown length
116465 118335: contig of 1865 bp in length
118336 118435: gap of unknown length
118435 120586: contig of 2151 bp in length
120586 120687: gap of unknown length
120687 121822: contig of 1136 bp in length
121822 121932: gap of unknown length
121932 124018: contig of 2056 bp in length
124018 124118: gap of unknown length
124119 125819: contig of 1701 bp in length
125820 125919: gap of unknown length
125920 127438: contig of 1519 bp in length
127439 127538: gap of unknown length
127539 129346: contig of 1808 bp in length
129347 129446: gap of unknown length
129446 132165: contig of 2719 bp in length
132165 132265: gap of unknown length
132266 134061: contig of 1796 bp in length
134062 134161: gap of unknown length
134162 136403: contig of 2242 bp in length
136404 136503: gap of unknown length
136504 138291: contig of 1788 bp in length
138292 138391: gap of unknown length
138392 140162: contig of 1771 bp in length
140163 140262: gap of unknown length
140263 142202: contig of 1940 bp in length
142203 142302: gap of unknown length
142303 144059: contig of 1757 bp in length
144060 144159: gap of unknown length
144160 145878: contig of 1719 bp in length
145879 145978: gap of unknown length
145979 147443: contig of 1465 bp in length

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 2033)
 Warren, C.E., Smokler, D.S., and Dennis, J.W.
 UDP-glucNAc:Gal beta1-3GalNAc-6 (GlcNAc to GalNAc) beta1-6GlcNAc
 transferase (core2-GnT) mouse cDNA sequence
 Unpublished
 2 (bases 1 to 2033)
 Warren, C.E.
 Direct Submission
 Submitted (29-DEC-1994) Charles E. Warren, Mount Sinai Hospital,
 Samuel Lunenfeld Research Institute, 600 University Avenue,
 Toronto, Ontario M5G 1X5, Canada
 Location/Qualifiers
 1..2033
 /organism="Mus musculus"
 /strain="DBA/2"
 /db_xref="taxon:10090"
 /cell_line="MDA-MB-231"
 1..2033
 /gene="C2-GnT"
 583..1869
 /gene="C2-GnT"
 /EC_number="2.4.1.102"
 /codon_start=1
 /evidence="experimental"
 /product="core2-GlcNAc-transferase"
 /protein_id="AA60948.1"
 /db_xref="GI:624294"
 /translation="MLPRLFRRLFSCPTTYFMVLVLSITSVLRIRKPEFVSF
 HLEAGDDPISNNCKTKILQDPBEIOKVALEILTVQKRPWPHDYINMRDCA
 FIKRKIVPELTKKEVEFPPIASIVVHNKIEMLDLRAIYMPQNYCIHVRKAE
 SFLLAOGVLAECDFNVASQLESVYASWTIRKADLNCMKDLYRMANMKYILNCG
 MDPPIKTNLEIVRKLKSGTENNLEKMPNKEERKRYAVVDGKLITGTIVKAP
 PLTPLESAYFVYREYVYVLENNIOKLEMAQDTYSPDELMATTQRIPEVPG
 SPSSNKYDLSDMNAIARFYKNQYFBEVDNGCARPYPCSGVHRVSVCFAGDLSMKL
 ROHHLFRANKFDMDVDPFALICLDEHLRRLKLENLEH"
 610..678
 /gene="C2-GnT"
 /note="encodes predicted Golgi transmembrane/signal
 sequence (non-cleaved)"
 754..762
 /gene="C2-GnT"
 /note="encodes potential N-glycosylation site (sequon)"
 865..876
 /gene="C2-GnT"
 /note="encodes potential N-glycosylation site (sequon)"
 misc_feature
 /note="encodes potential N-glycosylation site (sequon)"
 BASE COUNT 537 a 492 c 476 g 528 t
 ORIGIN
 Query Match 20.0%; Score 376.8; DB 10; Length 2033;
 Best Local Similarity 59.1%; Pred. No. 1.3e-108;
 Matches 645; Conservative 0; Mismatches 447; Indels 0; Gaps 0;

Db 1069 TTACCCGGGTCAGAGGCGATTGCATCCTGCTTTATATATGCTTTGGCCAGCCAGTTC 1128
 Qy 637 gtcgggtgtgtatgtccctcgtgtccagggtgcaagctgaccccaactgcatgaaagac 696
 Db 1129 GAGAGTGTGTTTATGGGTCCTCGACACGCGGTTAAAGACACACCTCACTGATGAAGAC 1188
 Qy 697 ttgctccagagctcagtgccgttgaataacttcctcgaatacattgtggagagcttcct 756
 Db 1189 CTGTACAGAAATGAATGCAAACTGAGAGTACTGTGATCATCTCTGTGATGATTTCCCT 1248
 Qy 757 ataaagacacatgcaagatgctcgaagctcgaagctcgaagctcgaagctcgaagctcga 816
 Db 1249 ATTAAACCAACCTGGAATTTGTCAGGAGCTCAAGTCTCCACAGGAGGAAACACCTG 1308
 Qy 817 gactcagagtgactcctcgaagacacaaagaaacccgctggaataatacattgtgagtgt 876
 Db 1309 GAACTGTGAAGATGCTCCCTCCCAACAGGAAGAAAGATGAAAGATGAGCGCTGTGTC 1368
 Qy 877 agagacacattac 936
 Db 1369 GATGGAGAGCTGACCAACACTGGGATGATGTCMAAGCAGCGCCCACTCAAAACTCTCTC 1428
 Qy 937 tttaagagaaatgctgacatgtgtgtgtcccgagatttgcacacatgtttgaagaac 996
 Db 1429 TTTTCAGCAGTGCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1488
 Qy 997 cctaaccac 1056
 Db 1489 GAAATATTTCAAAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1548
 Qy 1057 tgggac 1116
 Db 1549 TGGGCGACCACTCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1608
 Qy 1117 gacatcccaac 1176
 Db 1609 GACTGTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1668
 Qy 1177 atcgaataaggtgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1236
 Db 1669 GTTTCGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1728
 Qy 1237 tatggggtgggagcttgaattggaatggaatggaatggaatggaatggaatggaatgga 1296
 Db 1729 TTTGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1788
 Qy 1297 gaccacaaggtgagatgataatgataatgataatgataatgataatgataatgataatgata 1356
 Db 1789 GACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1848
 Qy 1357 atcattggaactggaacttggagacacacacacacacacacacacacacacacacacacac 1416
 Db 1849 CTGAGAGAACTTAAAGACTTAAGCGCTGTGACACCGCTGGGAGATTAAGATGCTCAAA 1908
 Qy 1417 atgtacaac 1476
 Db 1909 CTGTACTCTCAT 1920

Search completed: September 26, 2002, 04:26:52
 Job time: 9177 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 03:03:36 ; Search time 475.38 seconds
(without alignments)
6804.381 Million cell updates/sec

Title: US-09-874-390-1_COPY_436_2319
Perfect score: 1884
Sequence: 1 gatgtgtctctctccaccc.....ataataatacctaacaga 1884

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: N_Geneseq_032802.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1884	100.0	2319	21	Human C2/4Gnt CDNA
2	1833	97.3	2108	21	A core 2 beta-1,6-
3	1833	97.3	2236	22	Human colon cancer
4	1832.6	97.3	2229	21	Human pancreatic c
5	1798.8	95.5	2128	22	Nucleotide sequenc
6	1317	69.9	1317	21	DNA encoding a cor
7	1221	64.8	1221	21	Core 2 beta-1,6-N-
8	942.2	50.0	997	19	Human secreted pro
9	486.2	25.8	777	19	Human secreted pro

10	456	24.2	524	21	AAA96571	Noncoding region o
11	369.4	19.6	2204	19	AAK24043	Human core 2Gnt DN
12	369.4	19.6	2204	21	AAK25469	Human heart core 2
13	367.8	19.5	2105	15	AA061559	CDNA sequence of h
14	364.6	19.4	5010	19	AAK24042	Rat DH1 CDNA. Rat
15	364.6	19.4	5010	21	AAK24042	Diabetic rat heart
16	355	18.8	2102	18	AAK5468	Coding sequence fo
17	284.8	15.1	300	21	AAAT80112	Human colon cancer
18	199	10.6	997	22	AAAT1261	DNA encoding novel
19	191.8	10.2	1362	22	AA500045	Human DNA encoding
20	180.4	9.6	1807	16	AA089201	I-branching enzyme
21	180.4	9.6	1807	17	AAAT16201	Beta-1,6-N-acetylgl
22	180.4	9.6	1807	19	AAV30006	Full length CDNA s
23	180.4	9.6	1807	19	AAV16000	Human beta-1,6-N-a
24	162.8	8.6	378	17	AAAT16202	C2Gnt catalytic do
25	139.6	7.4	410	22	AA543357	DNA encoding novel
26	139.6	7.4	410	22	AAAT5732	Human polynucleoti
27	134	7.1	378	19	AAV30005	CDNA encoding a re
28	129.6	6.9	192	22	AAK62139	Human beta-1,6-N-
29	84	4.5	147	22	AAK62142	Murine beta-1,6-N-
30	69	3.7	69	21	AAK6572	Core 2 beta-1,6-N-
31	60	3.2	267	21	AAK6572	Noncoding region o
32	47.2	2.5	387	22	AAK64613	Novel human polynu
33	46.4	2.5	781	22	AAK98678	Human EST-derived
34	43.6	2.3	66	17	AAAT16204	C2Gnt C-terminal r
35	43	2.3	99	17	AAAT16203	C2Gnt C-terminal r
36	39.8	2.1	189	22	AAK65512	Human foetal liver
37	39.8	2.1	189	22	AAK33576	Probe #12042 for g
38	39.8	2.1	189	22	AAK14936	Human brain expres
39	39.8	2.1	189	22	AAK40659	Human bone marrow
40	39.8	2.1	189	22	AAK121432	Probe #11365 for g
41	39.8	2.1	189	22	AAK16721	Probe #15407 used
42	39.8	2.1	189	22	AAK107127	Probe #7118 used t
43	39.8	2.1	454	22	AAK53937	Human foetal liver
44	39.8	2.1	454	22	ABK33688	Probe #2154 for ge
45	39.8	2.1	454	22	AAK02207	Human brain expres

ALIGNMENTS

RESULT 1	AAA48623	standard; CDNA: 2319 BP.
ID	AAA48623	
XX	AAA48623	
AC	19-SEP-2000	(first entry)
XX		
DT	Human C2/4Gnt CDNA.	
DE		
XX	Human: C2/4Gnt; UDP-N-acetylglucosamine: O-glycan biosynthesis;	
KW	O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer; ss.	
KW		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	496..1812
FT		/*tag= a
FT	primer_bind	/product= "C2/4Gnt"
FT		586..605
FT	primer_bind	/*tag= b
FT		1794..1813
FT		/*tag= c
FT		2244..2249
FT	polyA_signal	/*tag= d
XX		
XX	WO200034449-A2.	
XX		
XX	15-JUN-2000.	
XX		
XX	03-DEC-1999;	99WO-DK00677.
XX		

Db 1996 atgctgctgttctctcaacccctaaacctagtagttcctccaactattctcaactagt 2055
Qy 1621 gagaatggaactgctgtgatagaggagatgaaggagatagtgtgtagaacttgat 1680
Db 2056 gagaatggaactgctgtgatagaggagatgaaggagatagtgtgtagaacttgat 2115
Qy 1661 ttcaagtgaatgctgctgtgttagctttccattctgtgagctgctgtccctaataatc 1740
Db 2116 ttcaagtgaatgctgctgtgttagctttccattctgtgagctgctgtccctaataatc 2175
Qy 1741 caggttgtagctgtgagggaggaacttgatgtaagaagaactccctctgtactgtt 1800
Db 2176 caggttgtagctgtgagggaggaacttgatgtaagaagaactccctctgtactgtt 2235
Qy 1801 aacttaaaataaagctccgactcaagaagtattacccttcttctgctcagatgccc 1860
Db 2236 aacttaaaataaagctccgactcaagaagtattacccttcttctgctcagatgccc 2295
Qy 1861 agaaataataataalctaaacaga 1884
Db 2296 agaaataataataalctaaacaga 2319

RESULT 2
AAA96570
ID AAA96570 standard; DNA; 2108 BP.

AC AAA96570;

DF 08-FEB-2001 (first entry)

DE A core 2 beta-1,6-N-acetylglucosaminyltransferase DNA fragment.

XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
XX cancer; cardiovascular disorder; inflammatory disorder; asthma;
XX rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
XX septic shock; adult respiratory distress syndrome; ARDS; cancer;
XX platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;
XX clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
XX diverticulitis; ulcerative colitis; ss.

OS Homo sapiens.

PN CA2296936-A1.

PD 03-AUG-2000.

PF 03-FEB-2000; 2000CA-2296936.

PR 03-FEB-1999; 99US-0118674.

PA (GLYC-) GLYCDESIGN INC.

PI Korczak B, Lew A;

DR WPI: 2000-594746/57.

XX New nucleic acid molecules of core 2

PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new

PT compositions for treatment of disorders mediated by the enzyme

PT including cancer, cardiovascular and inflammatory disorders.

PS Claim 4; Page 51-52; 66pp; English.

XX The present sequence encodes a partial human core 2

CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.

CC The polypeptide can be used to treat diseases and disorders, such as

CC cancer, cardiovascular disorders and inflammatory disorders including

CC asthma, rheumatoid arthritis, inflammatory bowel disease,

CC arteriosclerosis, septic shock, adult respiratory distress syndrome

CC (ARDS) and cancer. Various platelet-mediated pathologies such as

CC atherosclerosis and clotting can also be treated. The polypeptides of

CC the invention are predominantly expressed in gastrointestinal tissue

CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC Gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.

XX Sequence 2108 BP; 596 A; 464 C; 509 G; 539 T; 0 other;

Qy Query Match 97.3%; Score 1833; DB 21; Length 2108;

Db Best local similarity 100.0%; Pred. No. 0;

Matches 1833; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gattgtgctccctccctaccctccctgctgtctggttccactgttccattctgtgacg 60
Db 208 gattgtgctccctccctaccctccctgctgtctggttccactgttccattctgtgacg 267
Qy 61 atgttcaatggaagagactgtccagctgcatctattgttggtctgtgctgtatag 120
Db 268 atgttcaatggaagagactgtccagctgcatctattgttggtctgtgctgtatag 327
Qy 121 ctgtctgccaactgtgctctgtaaaccttcttcaagttgaagtgtgactctgacacttg 180
Db 328 ctgtctgccaactgtgctctgtaaaccttcttcaagttgaagtgtgactctgacacttg 387
Qy 181 ggtctggaagtcgaaggaatcttcaagccagctactgttaagataatctgtataattctgt 240
Db 388 ggtctggaagtcgaaggaatcttcaagccagctactgttaagataatctgtataattctgt 447
Qy 241 aaacttccagcaaaagagctctacaactgttcaaggggtccaccggaggaaccaagaagca 300
Db 448 aaacttccagcaaaagagctctacaactgttcaaggggtccaccggaggaaccaagaagca 507
Qy 301 gtgtctcagctatctgtaataacctgtgaaggttcaagaagaagcgagccttcaagac 360
Db 508 gtgtctcagctatctgtaataacctgtgaaggttcaagaagaagcgagccttcaagac 567
Qy 361 acccactaccctccctccctaccagagactgtgagccttcaaggggtcgaagaagaagtctata 420
Db 568 acccactaccctccctccctaccagagactgtgagccttcaaggggtcgaagaagaagtctata 627
Qy 421 cagttccactgagcaaaagaggtgaggttccctattgtcatctcatctgtagttat 480
Db 628 cagttccactgagcaaaagaggtgaggttccctattgtcatctcatctgtagttat 687
Qy 481 gagaagattgaaaaccttgaagaagctactgcgagctgtgtatgcccccaagaacataac 540
Db 688 gagaagattgaaaaccttgaagaagctactgcgagctgtgtatgcccccaagaacataac 747
Qy 541 tgtgtccatgtgagatgaagaagtcctccagaagaacttcaagaaggcggttaagaacatttt 600
Db 748 tgtgtccatgtgagatgaagaagtcctccagaagaacttcaagaaggcggttaagaacatttt 807
Qy 601 tctgtctcccaaatgtcttcatatagccagtaagctgtgtcggttattatgctccctcg 660
Db 808 tctgtctcccaaatgtcttcatatagccagtaagctgtgtcggttattatgctccctcg 867
Qy 661 tccaaggtgcaagctgacccaactgcatgtaagaagctgtgtccagagctcagtcggtg 720
Db 868 tccaaggtgcaagctgacccaactgcatgtaagaagctgtgtccagagctcagtcggtg 927
Qy 721 aaatacttccgtaataactgtggaagcgaacttctctataaagagaagaatgagatggtc 780
Db 928 aaatacttccgtaataactgtggaagcgaacttctctataaagagaagaatgagatggtc 987
Qy 781 caggttccaaagatgttaatggaaggaatagcatgagatgaagatgctccccaagaagac 840
Db 988 caggttccaaagatgttaatggaaggaatagcatgagatgaagatgctccccaagaagac 1047
Qy 841 aaagaaccgcgtggaataatcactttaggttagtgaagaagacacattacacctaacaac 900
Db 1047 aaagaaccgcgtggaataatcactttaggttagtgaagaagacacattacacctaacaac 1107

Db 1048 aaagaaccgctggaataatcacttgagtgtagagagacaattacacctaac 1107
QY 901 aagaagaagatctccctcccttaatttaactatgtttacaggaatgctacattgt 960
Db 1108 aagaagaagatctccctcccttaatttaactatgtttacaggaatgctacattgt 1167
QY 961 gcttcccgagatttcttccacaatgttttgaagaaccttaattccacaactgattga 1020
Db 1168 gcttcccgagatttcttccacaatgttttgaagaaccttaattccacaactgattga 1227
QY 1021 tgggttaaaagaccttaagcccaagatgaacacctctgggccaaccttcagctgacgg 1080
Db 1228 tgggttaaaagaccttaagcccaagatgaacacctctgggccaaccttcagctgacgg 1287
QY 1081 tggatgctgctctgttcccaaccaccacaagtaacatactcaagacatgacttatt 1140
Db 1288 tggatgctgctctgttcccaaccaccacaagtaacatactcaagacatgacttatt 1347
QY 1141 gccagagctgctcaagtgagcaggttcataagggagagacatgataaggtgctcttatt 1200
Db 1348 gccagagctgctcaagtgagcaggttcataagggagagacatgataaggtgctcttatt 1407
QY 1201 cccctgcttgatcccaacgagcagctatctgcttattgggctgggagcttgaattgg 1260
Db 1408 cccctgcttgatcccaacgagcagctatctgcttattgggctgggagcttgaattgg 1467
QY 1261 atgcttcaaaacatcacctctgttggccaacaagtttgacccaagtagatgaatgct 1320
Db 1468 atgcttcaaaacatcacctctgttggccaacaagtttgacccaagtagatgaatgct 1527
QY 1321 ctccagagcttagaagaatccctacgttatagaagccatctatgtagtgaactttgagac 1380
Db 1528 ctccagagcttagaagaatccctacgttatagaagccatctatgtagtgaactttgagac 1587
QY 1381 acaatgaagaagcttctcaccctggtgggcaagaagcactgaacaacatgctcagaactt 1440
Db 1588 acaatgaagaagcttctcaccctggtgggcaagaagcactgaacaacatgctcagaactt 1647
QY 1441 ctggagacagctggtgggagacagagcttgcaattcgttgcatcctttagataaga 1500
Db 1648 ctggagacagctggtgggagacagagcttgcaattcgttgcatcctttagataaga 1707
QY 1501 gggctgtatagaattggtgtagatgactcttgccttgcaaatgctgacctgggtga 1560
Db 1708 gggctgtatagaattggtgtagatgactcttgccttgcaaatgctgacctgggtga 1767
QY 1561 atgctgcttctcaccaccctcaacctagtagttcctcacaactatctcacaagt 1620
Db 1768 atgctgcttctcaccaccctcaacctagtagttcctcacaactatctcacaagt 1827
QY 1621 gagaatgaagaactgctgtgtagaggaagtgaaggaaggaatggtgtagagcaactgat 1680
Db 1828 gagaatgaagaactgctgtgtagaggaagtgaaggaaggaatggtgtagagcaactgat 1887
QY 1681 ttcagttgaatgctgctgtgtagatcttccattctgtgagcgcttccataattc 1740
Db 1888 ttcagttgaatgctgctgtgtagatcttccattctgtgagcgcttccataattc 1947
QY 1741 caggttgtagcgtgtagaggaagacttgaatggaagaagaaccttccctcttactggt 1800
Db 1948 caggttgtagcgtgtagaggaagacttgaatggaagaagaaccttccctcttactggt 2007
QY 1801 aacttaaaaaataagatccttgaatcaagta 1833
Db 2008 aacttaaaaaataagatccttgaatcaagta 2040

RESULT 3
AAH34463
ID AAH34463 standard; cDNA; 2236 BP.
XX
AC AAH34463;
XX

DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1545.
XX
KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; chromosome 15; 5S.
XX
OS Homo sapiens.
XX
PN W0200122920-A2.
XX
PD 05-APR-2001.
XX
PE 28-SEP-2000; 2000MO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
XX
DR P-PSDB; AAG75058.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 1; Page 3190-3191; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 2236 BP; 609 A; 503 C; 548 G; 574 T; 2 other;

Query Match 97.3%; Score 1833; DB 22; Length 2236;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
Matches 1833; Conservative 0;

QY 1 gattgtgctcctccacccttccctgtgctggtctccaccctgtccaccattgtgagc 60
Db 374 gattgtgctcctccacccttccctgtgctggtctccaccctgtccaccattgtgagc 433
QY 61 atggttcaatggaagaactctgcagctgcatctactgttggctctgggctgctatag 120
Db 434 atggttcaatggaagaactctgcagctgcatctactgttggctctgggctgctatag 493
QY 121 ctgctggcaccctgtgcttgaacttctttaggttgaagtgaactctgacccctg 180
Db 494 ctgctggcaccctgtgcttgaacttctttaggttgaagtgaactctgacccctg 553
QY 181 ggtctggaatcaggaatctcaagccagctactgtagagatatctgtataattctcgt 240
Db 554 ggtctggaatcaggaatctcaagccagctactgtagagatatctgtataattctcgt 613

QY 241 aaacttcagcaagaagtgctatactatgttcagggtccaccgagggagccaaagaca 300
|||||
Db 614 aaacttcagcaagaagtgctatactatgttcagggtccaccgagggagccaaagaca 673
QY 301 gtgtctcagggtatctctgaataaccttgaggtcgaagaagcgagagctttccagac 360
|||||
Db 674 ggtcttcagggtatctctgaataaccttgaggtcgaagaagcgagagctttccagac 733
QY 361 acccctcctctccctccacccagagactgtgagcacttcaagcgctgaagaaggttcata 420
|||||
Db 734 acccctcctctccctccacccagagactgtgagcacttcaagcgctgaagaaggttcata 793
QY 421 cagttccacctgagcaaaagaggtgaggtccctatcgcatcctatctgtgtatcat 480
|||||
Db 794 cagttccacctgagcaaaagaggtgaggtccctatcgcatcctatctgtgtatcat 853
QY 481 gagaagattgaaaccttgaaaagctactcgagctgtgtatgtccctcagaacatatac 540
|||||
Db 854 gagaagattgaaaccttgaaaagctactcgagctgtgtatgtccctcagaacatatac 913
QY 541 tgtgtccatgtgagtagaagaatcccccagaaccttccaagaagcggtccaagaacttat 600
|||||
Db 914 tgtgtccatgtgagtagaagaatcccccagaaccttccaagaagcggtccaagaacttat 973
QY 601 tctgtctcccaaatgtctctcaataagctgtgtcgggtgtttatctctctcg 660
|||||
Db 974 tctgtctcccaaatgtctctcaataagctgtgtcgggtgtttatctctctcg 1033
QY 661 tccaggtgtcgaagctggaactccaactgtcatggaagactgtctccaagagctcagtcggtg 720
|||||
Db 1034 tccaggtgtcgaagctggaactccaactgtcatggaagactgtctccaagagctcagtcggtg 1093
QY 721 aaatctctctgaataacatggtggaagcttccctataaagaagcaatgcaagatgtgc 780
|||||
Db 1094 aaatctctctgaataacatggtggaagcttccctataaagaagcaatgcaagatgtgc 1153
QY 781 caggtctccaagaatgtctgaatggaagaatagcatgagtcagaggttacctcctaagcac 840
|||||
Db 1154 caggtctccaagaatgtctgaatggaagaatagcatgagtcagaggttacctcctaagcac 1213
QY 841 aaagaaccgctgtgaaataatacaacttgaagtgtagagaagacacattacctaaccac 900
|||||
Db 1214 aaagaaccgctgtgaaataatacaacttgaagtgtagagaagacacattacctaaccac 1273
QY 901 aagaagaagaatcccccctataattcaactatgtttacaggggaatgctacatgtgtg 960
|||||
Db 1274 aagaagaagaatcccccctataattcaactatgtttacaggggaatgctacatgtgtg 1333
QY 961 gcttcccgagattcgtccacaacatgttttgaagaaccttaaaatcccaacaactgtatga 1020
|||||
Db 1334 gcttcccgagattcgtccacaacatgttttgaagaaccttaaaatcccaacaactgtatga 1393
QY 1021 tgggttaaaagacattatagccagatgaacaccttgggccaacctttagagtgtaggg 1080
|||||
Db 1394 tgggttaaaagacattatagccagatgaacaccttgggccaacctttagagtgtaggg 1453
QY 1081 tggatgcctggtctgttcccaaccaccccaagtaagacatctcagacatgacttatct 1140
|||||
Db 1454 tggatgcctggtctgttcccaaccaccccaagtaagacatctcagacatgacttatct 1513
QY 1141 gccaggtctggtcaagtgtgcaaggtcctatgtgttaagggctgtgggaacttgatgtg 1200
|||||
Db 1514 gccaggtctggtcaagtgtgcaaggtcctatgtgttaagggctgtgggaacttgatgtg 1573
QY 1201 cctgtctcgtgaatccaccagcggtatctgtgttaagggctgtgggaacttgatgtg 1260
|||||
Db 1574 cctgtctcgtgaatccaccagcggtatctgtgttaagggctgtgggaacttgatgtg 1633
QY 1261 atgcttcaaaacacatcactgtgtgccaacaagtttgacccaagaagtatgataatgtct 1320
|||||
Db 1634 atgcttcaaaacacatcactgtgtgccaacaagtttgacccaagaagtatgataatgtct 1693
QY 1321 ctctcagtgcttagaagaataactacgttataagggcatctatgtgagctttagagac 1380

|||||
Db 1694 ctctcagtgcttagaagaataactacgttataagggcatctatgtggagactttagagac 1753
QY 1381 acactatgagagcggtgtctactcgttggggcaagagcatgtataacaatgtctagaactg 1440
|||||
Db 1754 acactatgagagcggtgtctactcgttggggcaagagcatgtataacaatgtctagaactg 1813
QY 1441 ctgggacaggtgtgggttgggaaccagggcttggcaattcgtggcatctttaggataga 1500
|||||
Db 1814 ctgggacaggtgtgggttgggaaccagggcttggcaattcgtggcatctttaggataga 1873
QY 1501 gggctgtctatagatgttgggtgaatagatcttctgcttgcgaattgctgctgtgggtga 1560
|||||
Db 1874 gggctgtctatagatgttgggtgaatagatcttctgcttgcgaattgctgctgtgggtga 1933
QY 1561 atgtgtctgttctctcaaccctcaaccctagtagttcctcctaacttctcctaagt 1620
|||||
Db 1934 atgtgtctgttctctcaaccctcaaccctagtagttcctcctaacttctcctaagt 1993
QY 1621 gagaatgagaactgtctgtatagaggaggtgaaggagagatatgtgttagagcacttgat 1680
|||||
Db 1994 gagaatgagaactgtctgtatagaggaggtgaaggagagatatgtgttagagcacttgat 2053
QY 1681 ttcagttgaaatgcctgtgtgtatgacttccatctctgttggagctgtccgttccaataatc 1740
|||||
Db 2054 ttcagttgaaatgcctgtgtgtatgacttccatctctgttggagctgtccgttccaataatc 2113
QY 1741 caggttggtagcgttggagagaaacttgaatggaagaagaaaccttccctctgtactgt 1800
|||||
Db 2114 caggttggtagcgttggagagaaacttgaatggaagaagaaaccttccctctgtactgt 2173
QY 1801 aacttaaaataaataagctcctgattccaagta 1833
|||||
Db 2174 aacttaaaataaataagctcctgattccaagta 2206

RESULT 4
AAC99109
ID AAC99109 standard; cDNA; 2229 BP.
XX
AC AAC99109;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:337.
XX
OS Homo sapiens.
XX
PN WO200055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-579444/54.
XX
DR P-PSDB: AAB54344.
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
treating, or ameliorating a medical condition, particular pancreatic

PT cancer, or for use in assays for diagnosing a pathological condition -
XX Claim 1: Page 759-760; 1379pp; English.
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, nontropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 2229 BP; 605 A; 501 C; 546 G; 572 T; 5 other;

Query Match 97.3%; Score 1832.6; DB 21; Length 2229;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1832; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 gatgtgtccctcccaacctccctgtgctgcggtcctcacctgtcccttctgtgacg 60
Db |||||||
QY 367 gatgtgtcctctcccaacctccctgtgctgcggtcctcacctgtcccttctgtgacg 426
Db |||||||
QY 61 atggttcaatggaagagactctgcacgtcgtacattgttgggtctggtcgtgataatg 120
Db 427 atggttcaatggaagagactctgcacgtcgtacattgttgggtctggtcgtgataatg 486
QY 121 ctgtgtgccaactgtggtctggaacttcttcaaggttgaagtgtgaactctgacacttg 180
Db |||||||
QY 487 ctgtgtgccaactgtggtctggaacttcttcaaggttgaagtgtgaactctgacacttg 546
Db |||||||
QY 181 ggtctggaagccaagagatctcaagccaagtactgtagaatacttgtaattctctg 240
Db 547 ggtctggaagccaagagatctcaagccaagtactgtagaatacttgtaattctctg 606
QY 241 aaacttcagcaaaagaggtctatcaactgttcaagggttcacccgaggggaccaaagagca 300
Db 607 aaacttcagcaaaagaggtctatcaactgttcaagggttcacccgaggggaccaaagagca 666
QY 607 aaacttcagcaaaagaggtctatcaactgttcaagggttcacccgaggggaccaaagagca 666
QY 301 gtgttcaagctatctcgaataaccttgaggttgaagaagaagcgagagcctttcacagac 360
Db 667 gtgttcaagctatctcgaataaccttgaggttgaagaagaagcgagagcctttcacagac 726
QY 361 acccaactactctccctcacacagagactgtgagcaacttcaaggttgaagaagaattcata 420
Db 727 acccaactactctccctcacacagagactgtgagcaacttcaaggttgaagaagaattcata 786
QY 421 cagttccactgagcaaaagaggtgaggttccctattgcatatactatgtgtatcat 480
Db 787 cagttccactgagcaaaagaggtgaggttccctattgcatatactatgtgtatcat 846
QY 481 gagaagaattgaanaacttgaagaagctactcgagagctgttatagcccctcagaataatcac 540
Db 847 gagaagaattgaanaacttgaagaagctactcgagagctgttatagcccctcagaataatcac 906
QY 541 tgttccatgtgatagtgaaggtccccaagaacttcaagaaggggtcagaagaattatc 600
Db 907 tgttccatgtgatagtgaaggtccccaagaacttcaagaaggggtcagaagaattatc 966

QY 601 tcttgtctcccaaatgtcttcatagccagtaagctgttccggttggttattcctctcgtg 660
Db |||||||
QY 967 tcttgtctcccaaatgtcttcatagccagtaagctgttccggttggttattcctctcgtg 1026
Db |||||||
QY 661 tcaaggttgaagaactgtgacccaactgtgataaggaagactgttcccaaggtcaggtgcgtg 720
Db 1027 tcaaggttgaagaactgtgacccaactgtgataaggaagactgttcccaaggtcaggtgcgtg 1086
QY 721 aaatactctgtataaatgttggagcagacttccctataagaagcaatgagagatgtgc 780
Db 1087 aaatactctgtataaatgttggagcagacttccctataagaagcaatgagagatgtgc 1146
QY 781 caggcttcaagaatgttgaatgggaggaatagacatgagtcagaggtaccccttaagcac 840
Db 1147 caggcttcaagaatgttgaatgggaggaatagacatgagtcagaggtaccccttaagcac 1206
QY 841 aagaagaacccgctgggaataatcacttgaagtgtgagagacatcattacccaaccaac 900
Db 1207 aagaagaacccgctgggaataatcacttgaagtgtgagagacatcattacccaaccaac 1266
QY 901 aagaagaagatccctcccttaataattacatgttcaagaagatgtgacatgtg 960
Db 1267 aagaagaagatccctcccttaataattacatgttcaagaagatgtgacatgtg 1326
QY 961 gttcccgagattcgttccacaatgttgaagaacccaataatccacaacatgtatgaa 1020
Db 1327 gttcccgagattcgttccacaatgttgaagaacccaataatccacaacatgtatgaa 1386
QY 1021 tgggttaaaagacattatgccaagatgaacacctcctgggccaaccttaagcgtgacag 1080
Db 1447 tgggttaaaagacattatgccaagatgaacacctcctgggccaaccttaagcgtgacag 1506
QY 1141 gccaggtctgccaagatgagcaggttcatbaggaggaacatcgataaagggtcctctatgct 1200
Db 1507 gccaggtctgccaagatgagcaggttcatbaggaggaacatcgataaagggtcctctatgct 1566
QY 1201 cccctgctctggaatcccaacgagcgtatctggttatgttggtcgtggagcttgaatttg 1260
Db 1567 cccctgctctggaatcccaacgagcgtatctggttatgttggtcgtggagcttgaatttg 1626
QY 1261 atgcttcaaaaacacacacccgttggccaacaagtcttgaccaagaagttagatgaatgct 1320
Db 1627 atgcttcaaaaacacacacccgttggccaacaagtcttgaccaagaagttagatgaatgct 1686
QY 1321 ctccaaggtctagaagaatactcaagcttataagggcatctatgtggagactttagagac 1380
Db 1687 ctccaaggtctagaagaatactcaagcttataagggcatctatgtggagactttagagac 1746
QY 1381 acactatgagaagcgttgcctacccctgttgggccaagaagcatgtacaacaatgtcagaacttg 1440
Db 1747 acactatgagaagcgttgcctacccctgttgggccaagaagcatgtacaacaatgtcagaacttg 1806
QY 1441 ctgggaacagttgtgtgaggaacagaggtcttgcaattctgtgacatccctttagaataaga 1500
Db 1807 ctgggaacagttgtgtgaggaacagaggtcttgcaattctgtgacatccctttagaataaga 1866
QY 1501 gggctgctatataatgttgggtatagatccttggccttgcaaatgtcgtcgtgtgtga 1560
Db 1867 gggctgctatataatgttgggtatagatccttggccttgcaaatgtcgtcgtgtgtga 1926
QY 1561 atgctgtgttcttccacccctcaaaccttagtagttccctcaactaacttctcaagaagt 1620
Db 1927 atgctgtgttcttccacccctcaaaccttagtagttccctcaactaacttctcaagaagt 1986
QY 1621 gagaatbagaagactgctgtgaataggaagatgaaagaggaatacgtgtgataagacattgat 1680
Db 1987 gagaatbagaagactgctgtgaataggaagatgaaagaggaatacgtgtgataagacattgat 2046


```

QY 1081 tggatgctggtctctgttcccaaccacccaagtaacacatcctcaagacatgactctatt 1140
    |||
Db 1374 tggatgctggtctctgttcccaaccacccaagtaacacatcctcaagacatgactctatt 1433
QY 1141 gccaggtctgtaagtgagcaggttcataagagagacatgataaggtgtctcttaagct 1200
    |||
Db 1434 gccaggtctgtaagtgagcaggttcataagagagacatgataaggtgtctcttaagct 1493
QY 1201 ccctgctctgtaatccacccagcgactatctgtctatagggtggcctggaattgg 1260
    |||
Db 1494 ccctgctctgtaatccacccagcgactatctgtctatagggtggcctggaattgg 1353
QY 1261 atgcttcaaaacacatccactgtctgagccacaagtctgacccaaggtagaatgaatgct 1320
    |||
Db 1554 atgcttcaaaacacatccactgtctgagccacaagtctgacccaaggtagaatgaatgct 1613
QY 1321 ctccagtgcttaagaagaatacctacgttataaaggccatctatggaacttgagac 1380
    |||
Db 1614 ctccagtgcttaagaagaatacctacgttataaaggccatctatggaacttgagac 1673
QY 1381 acacatgagagcgttgctacccctgagggaagacatgatacaaacatgctcagaattg 1440
    |||
Db 1674 acacatgagagcgttgctacccctgagggaagacatgatacaaacatgctcagaattg 1733
QY 1441 ctggagacagtgtggtgaggaacacaggcttgcacatcgtggacatccttaagataaga 1500
    |||
Db 1734 ctggagacagtgtggtgaggaacacaggcttgcacatcgtggacatccttaagataaga 1793
QY 1501 gggctgtctatagaatgttgagtaagtagatcctttgctcgaacatgctgcctgggtga 1560
    |||
Db 1794 gggctgtctatagaatgttgagtaagtagatcctttgctcgaacatgctgcctgggtga 1853
QY 1561 atgctgtctgttctctacacccctaacacctagtagttccctcactaaccttctcactaagt 1620
    |||
Db 1854 atgctgtctgttctctacacccctaacacctagtagttccctcactaaccttctcactaagt 1913
QY 1621 gagaatgagacatcgtctgtatagggagagatgaaaggagatatgtgttaagacacttgat 1680
    |||
Db 1914 gagaatgagacatcgtctgtatagggagagatgaaaggagatatgtgttaagacacttgat 1973
QY 1681 ttcaagttgaatgcctgctgtgtaagcttttcattctctgagcgccgcttcataaattc 1740
    |||
Db 1974 ttcaagttgaatgcctgctgtgtaagcttttcattctctgagcgccgcttcataaattc 2033
QY 1741 caggtttgtagcgttgagagagaacttgatggaagaagacatccctctgtactgctt 1800
    |||
Db 2034 caggtttgtagcgttgagagagaacttgatggaagaagacatccctctgtactgctt 2091
QY 1801 aacttaaaataata 1816
    |||
Db 2092 aacttaaaataata 2107
    |||

RESULT 6
ID AAA96569 standard: DNA: 1317 BP.
XX AAA96569;
AC AAA96569;
DT 08-FEB-2001 (first entry)
DE DNA encoding a core 2 beta-1,6-N-acetylglucosaminyltransferase.
XX
XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GLCNAC-T;
KW cancer; cardiovascular disorder; inflammatory bowel disease; arteriosclerosis;
KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis; ss.
XX
XX Homo sapiens.
OS

```

```

XX FH Key Location/Qualifiers
XX FT CDS 1..1317
XX FT /tag= a
XX FT /product= "core 2
FT beta-1,6-N-acetylglucosaminyltransferase"
PN CA2296936-A1.
XX
XX 03-AUG-2000.
PD
XX
XX 03-FEB-2000; 2000CA-2296936.
PF
XX
XX 03-FEB-1999; 99US-0118674.
PR
XX
XX (GLYC-) GLYCODESIGN INC.
PA
XX
XX Korceak B, Lew A;
PI
XX
XX WPI; 2000-594746/57.
DR
XX
XX P-PSDB; AAB18995.
DR
XX
XX New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
XX Claim 4; Page 50; 66pp; English.
PS
XX
XX The present sequence encodes a human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GLCNAC-T) polypeptide.
CC The polypeptide can be used to treat diseases and disorders, such as
CC cancer, cardiovascular disorders and inflammatory disorders including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC Gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
XX Sequence 1317 BP; 359 A; 307 C; 322 G; 329 T; 0 other;
SO

```

Query Match 69.9%; Score 1317; DB 21; Length 1317;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 61 atggttcaatggaagactctgcagctgcaatcttggccttggtgactatg 120
    |||
Db 1 atggttcaatggaagactctgcagctgcaatcttggccttggtgactatg 60
QY 121 ctgctggcactgtggtctgaaacttcttcaagttgaagtgtgactgaccactg 180
    |||
Db 61 ctgctggcactgtggtctgaaacttcttcaagttgaagtgtgactgaccactg 120
QY 181 ggtctggaatccaggaatccaaagccagctactgtagaataatctgtataattctcg 240
    |||
Db 121 ggtctggaatccaggaatccaaagccagctactgtagaataatctgtataattctcg 180
QY 241 aaacttcagcaagaggtctacactgttcaagggtacaccgagggagccaagagga 300
    |||
Db 181 aaacttcagcaagaggtctacactgttcaagggtacaccgagggagccaagagga 240
QY 301 gtgcttcaagctattctgataaactgtgaggttcaagaagaagcgagagcttcaagac 360
    |||
Db 241 gtgcttcaagctattctgataaactgtgaggttcaagaagaagcgagagcttcaagac 300

```

QY 361 accactactctctccctccaccagagactgtgagcacttcaagctgtaagaagtcatata 420
DB 301 accactactactctccctccaccagagactgtgagcacttcaagctgtaagaagtcatata 360
QY 421 cagttccactgtgacaaagaaggtgaggttcctctattgcatctctatgttgattcat 480
DB 361 cagttccactgtgacaaagaaggtgaggttcctctattgcatctctatgttgattcat 420
QY 481 gagaagatgtgaactttgaaaggctactgtgagctgtgtatgtgcccccaaaatataac 540
DB 421 gagaagatgtgaactttgaaaggctactgtgagctgtgtatgtgcccccaaaatataac 480
QY 541 tgtgtccatgtgtgagaaagtccccaagaacttccaagaagggtgtcaagaactatatt 600
DB 481 tgtgtccatgtgtgagaaagtccccaagaacttccaagaagggtgtcaagaactatatt 540
QY 601 tcttgcttcccaaatgttcttcatagccaagtgtgtcggtgtgtttatgtctctctg 660
DB 541 tcttgcttcccaaatgttcttcatagccaagtgtgtcggtgtgtttatgtctctctg 600
QY 661 tccaggggtgacaaagtgtgacactgtgacaaagtgtgacaaagtgtgacaaagtgtg 720
DB 601 tccaggggtgacaaagtgtgacactgtgacaaagtgtgacaaagtgtgacaaagtgtg 660
QY 721 aaatacttctgaaatatacatgtgtgagacgacttctctataaagaagcaatgtcagatgtc 780
DB 661 aaatacttctgaaatatacatgtgtgagacgacttctctataaagaagcaatgtcagatgtc 720
QY 781 caggtcttcaagatgttgaatgtgagagaaatagcatgtgagatcaagatgtcctctaagcac 840
DB 721 caggtcttcaagatgttgaatgtgagagaaatagcatgtgagatcaagatgtcctctaagcac 780
QY 841 aaaaaaacccgctgtgaatatcatctgtttaggttagtgagagacacattacataaccacac 900
DB 781 aaaaaaacccgctgtgaatatcatctgtttaggttagtgagagacacattacataaccacac 840
QY 901 aagaagaagaatctctccctctataatctaactgtttaaaggaaggaatgtctacattgtg 960
DB 841 aagaagaagaatctctccctctataatctaactgtttaaaggaaggaatgtctacattgtg 900
QY 961 gcttcccgagatctcgttccaacatgtttaaagaacccctaaatcccaacacgtatgttaa 1020
DB 901 gcttcccgagatctcgttccaacatgtttaaagaacccctaaatcccaacacgtatgttaa 960
QY 1021 tgggttaaaagacacttataagccagatgaaacactctgtggccacccttcaacgtgtgacgg 1080
DB 961 tgggttaaaagacacttataagccagatgaaacactctgtggccacccttcaacgtgtgacgg 1020
QY 1081 tggatgctgtgtctgttcccaacaccccaagtaagacatctcagacatgtactatatt 1140
DB 1021 tggatgctgtgtctgttcccaacaccccaagtaagacatctcagacatgtactatatt 1080
QY 1141 gccagagctgtgtcgaagtgtgacaggtatcatgtaggtgagacatcgaataaggtgtcctatgt 1200
DB 1081 gccagagctgtgtcgaagtgtgacaggtatcatgtaggtgagacatcgaataaggtgtcctatgt 1140
QY 1201 cccgtgtctgtgaatcccaacacgggtgtatctgctgttatagggtgtgagacttgaattgtg 1260
DB 1141 cccgtgtctgtgaatcccaacacgggtgtatctgctgttatagggtgtgagacttgaattgtg 1200
QY 1261 atgtcttcaaaacacacacttctgtgccaacaagtgtgacccaagaagtgtgagataatgt 1320
DB 1201 atgtcttcaaaacacacacttctgtgccaacaagtgtgacccaagaagtgtgagataatgt 1260
QY 1321 cttaagtgcttagaagaatatacctacgttataaaggccatctatgtgagactgtgaacttga 1377
DB 1261 cttaagtgcttagaagaatatacctacgttataaaggccatctatgtgagactgtgaacttga 1321

RESULT 7
AAA96575
ID AAA96575 standard; DNA; 1221 BP.
XX

AC AAA96575;
XX
DT 08-FEB-2001 (first entry)
DE Core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region DNA.
XX
DE Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
XX
KW cancer; cardiovascular disorder; inflammatory bowel disease; arteriosclerosis;
KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
KW platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;
KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis; ss.
XX
OS Homo sapiens.
XX
PN CA2296936-A1.
XX
PD 03-AUG-2000.
XX
PE 03-FEB-2000; 2000CA-2296936.
XX
PR 03-FEB-1999; 99US-0118674.
XX
PA (GLYC-) GLYCDESIGN INC.
XX
PI Korczak B, Lew A;
XX
DR WPI: 2000-594746/57.
XX
PT New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
PS Claim 4; Page 54; 66pp; English.
XX
CC The present sequence encodes a fragment of a human core 2
CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
CC The polypeptide can be used to treat diseases and disorders, such as
CC cancer, cardiovascular disorders and inflammatory disorders including
CC asthma, rheumatoid arthritis, inflammatory bowel disease,
CC arteriosclerosis, septic shock, adult respiratory distress syndrome
CC (ARDS) and cancer. Various platelet-mediated pathologies such as
CC atherosclerosis and clotting can also be treated. The polypeptides of
CC the invention are predominantly expressed in gastrointestinal tissue
CC (stomach, colon, intestine, testis) and are elevated in cancer.
CC Gastrointestinal disorders that may be prevented or treated include
CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
CC and ulcerative colitis. The antibodies may be used in
CC immuno-histochemical analysis, to detect the novel polypeptide and to
CC localize it to particular cells and tissues and to specific subcellular
CC locations and to quantitate the level of expression.
XX
SQ Sequence 1221 BP; 342 A; 285 C; 295 G; 299 T; 0 other;

Query Match 64.8%; Score 1221; DB 21; Length 1221;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1221; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 ttgaagtgtagcttgaccacttggtgtgaggtcccaaggaatcccaagcagtagt 216
DB 1 ttgaagtgtagcttgaccacttggtgtgaggtcccaaggaatcccaagcagtagt 60
QY 217 aggaatatctgtataattctctgaacttcccaagaagaggtctatataactgttcagg 276
DB 61 aggaatatctgtataattctctgaacttcccaagaagaggtctatataactgttcagg 120
QY 277 gtacccgagggagacaaagagcagtgcttcaagctatctgaataacctgtgagtcagg 336
DB 121 gtacccgagggagacaaagagcagtgcttcaagctatctgaataacctgtgagtcagg 180

QY	337	aagaagagagagagcctttccaaagacaccacactactcctccctccaaagaaactgtagaac	396
Dp	181	aagaagagagagagcctttccaaagacaccacactactcctccctccaaagaaactgtagaac	240
QY	397	ttcaagagctgaaaggaagtgtcatatagctttcccactgagcaaaagaggtgtgattccct	456
Dp	241	ttcaagagctgaaaggaagtgtcatatagctttcccactgagcaaaagaggtgtgattccct	300
QY	457	attgcatacctcttbtggtattctcatgagagatgtgaaacttggaaaggtactgtgaact	516
Dp	301	attgcatacctcttbtggtattctcatgagagatgtgaaacttggaaaggtactgtgaact	360
QY	517	gtgtagtccccctcagaacaatactgtgtccatgtgtgaagagaggtccccaagaacttc	576
Dp	361	gtgtagtccccctcagaacaatactgtgtccatgtgtgaagagaggtccccaagaacttc	420
QY	577	aaagaagcggttcaaaagcaatattcttctgtcttcccaaatgtcttcatagcagtagctg	636
Dp	421	aaagaagcggttcaaaagcaatattcttctgtcttcccaaatgtcttcatagcagtagctg	480
QY	637	gttcgggttggtttatgtcctctctgtccaaagtgtcaagctgactccaactgcatgtgaagac	696
Dp	481	gttcgggttggtttatgtcctctctgtccaaagtgtcaagctgactccaactgcatgtgaagac	540
QY	697	ttgtctccagagctcgaatgtccgtgtgaaatactctctgtaatacatgtgtgagacgattccct	756
Dp	541	ttgtctccagagctcgaatgtccgtgtgaaatactctctgtaatacatgtgtgagacgattccct	600
QY	757	ataaagagcaatbtcaagagatgtgtccaaagctccctcaaaagtgttaaaagtggagagatagcag	816
Dp	601	ataaagagcaatbtcaagagatgtgtccaaagctccctcaaaagtgtgtaaagtggagagatagcag	660
QY	817	gagtcagaaggttacctccttaagcacaagaagaaccggtgtgaaatatacatctttagagtagtg	876
Dp	661	gagtcagaaggttacctccttaagcacaagaagaaccggtgtgaaatatacatctttagagtagtg	720
QY	877	agagaacacattacaacctcaaaccaagaagaagatctccctcccttatatttaactatg	936
Dp	721	agagaacacattacaacctcaaaccaagaagaagatctccctcccttatatttaactatg	780
QY	937	tttccaaaggaatbtgtaacatctgtgagcttcccgagatcttcgcacaagaatgttttgaagaac	996
Dp	781	tttccaaaggaatbtgtaacatctgtgagcttcccgagatcttcgcacaagaatgttttgaagaac	840
QY	997	ccttaaatccccaacacactgattgaatgtgtgtaaaagaacattataagcccgatgaacacctc	1056
Dp	841	ccttaaatccccaacacactgattgaatgtgtgtaaaagaacattataagcccgatgaacacctc	900
QY	1057	tggcgcaaccttcagcgtgtcagcagctgtgaatgtgcctgtgcctgttcccaaccaaccacaagtac	1116
Dp	901	tggcgcaaccttcagcgtgtcagcagctgtgaatgtgcctgtgcctgttcccaaccaaccacaagtac	960
QY	1117	gacatctcagacatgactcttattctccaaagctgtgtcagaagtgtgcagaaggttcatatgagggagac	1176
Dp	961	gacatctcagacatgactcttattctccaaagctgtgtcagaagtgtgcagaaggttcatatgagggagac	1020
QY	1177	atcgataaaggtgtcctcttattgtcttcctctcctcctcggaatcccaacagcgggtctatccgtgct	1236
Dp	1021	atcgataaaggtgtcctcttattgtcttcctctcctcctcggaatcccaacagcgggtctatccgtgct	1080
QY	1237	tatggggctgtgggacttgaatttgaatgtgttcaaaacacatcacactgttggccaacaagttt	1296
Dp	1081	tatggggctgtgggacttgaatttgaatgtgttcaaaacacatcacactgttggccaacaagttt	1140
QY	1297	gacccaaaagtagatgtataatgtccttcagtgctttagaagaaatacctaagtataaagcc	1356
Dp	1141	gacccaaaagtagatgtataatgtccttcagtgctttagaagaaatacctaagtataaagcc	1200
QY	1357	atctcatgtgagctgaactttga 1377	
Dp	1201	atctcatgtgagctgaactttga 1221	

RESULT	ID	AAV59800	standard; DNA; 997 BP.
8	AAV59800;		
19-JAN-1999	(first entry)		
Human secreted protein gene 171 clone HTEDJ34.			
Human; secreted protein; fusion protein; gene therapy; protein therapy; diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia; developmental abnormality; foetal deficiency; blood; allergy; renal; ds; immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma; inflammation; ischaemic shock; Alzheimer's disease; retinosis; AIDS; cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus; osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion; endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.			
Homo sapiens.			
WO9839448-A2.			
11-SEP-1998.			
06-MAR-1998;	98WO-US04493.		
02-OCT-1997;	97US-0061060.		
07-MAR-1997;	97US-0038621.		
07-MAR-1997;	97US-0040161.		
07-MAR-1997;	97US-0040162.		
07-MAR-1997;	97US-0040163.		
07-MAR-1997;	97US-0040333.		
07-MAR-1997;	97US-0040334.		
07-MAR-1997;	97US-0040336.		
07-MAR-1997;	97US-0040626.		
11-APR-1997;	97US-0043311.		
11-APR-1997;	97US-0043312.		
11-APR-1997;	97US-0043313.		
11-APR-1997;	97US-0043314.		
11-APR-1997;	97US-0043568.		
11-APR-1997;	97US-0043569.		
11-APR-1997;	97US-0043576.		
11-APR-1997;	97US-0043580.		
11-APR-1997;	97US-0043580.		
11-APR-1997;	97US-0043580.		
11-APR-1997;	97US-0043670.		
11-APR-1997;	97US-0043671.		
11-APR-1997;	97US-0043672.		
11-APR-1997;	97US-0043674.		
23-MAY-1997;	97US-0047492.		
23-MAY-1997;	97US-0047500.		
23-MAY-1997;	97US-0047501.		
23-MAY-1997;	97US-0047502.		
23-MAY-1997;	97US-0047503.		
23-MAY-1997;	97US-0047581.		
23-MAY-1997;	97US-0047582.		
23-MAY-1997;	97US-0047583.		
23-MAY-1997;	97US-0047584.		
23-MAY-1997;	97US-0047585.		
23-MAY-1997;	97US-0047586.		
23-MAY-1997;	97US-0047587.		
23-MAY-1997;	97US-0047588.		
23-MAY-1997;	97US-0047589.		
23-MAY-1997;	97US-0047590.		
23-MAY-1997;	97US-0047592.		
23-MAY-1997;	97US-0047593.		
23-MAY-1997;	97US-0047594.		
23-MAY-1997;	97US-0047595.		
23-MAY-1997;	97US-0047596.		
23-MAY-1997;	97US-0047597.		
23-MAY-1997;	97US-0047598.		
23-MAY-1997;	97US-0047599.		

PR 23-MAY-1997; 97US-0047600.
 PR 23-MAY-1997; 97US-0047601.
 PR 23-MAY-1997; 97US-0047602.
 PR 23-MAY-1997; 97US-0047613.
 PR 23-MAY-1997; 97US-0047614.
 PR 23-MAY-1997; 97US-0047615.
 PR 23-MAY-1997; 97US-0047617.
 PR 23-MAY-1997; 97US-0047618.
 PR 23-MAY-1997; 97US-0047619.
 PR 23-MAY-1997; 97US-0047623.
 PR 06-JUN-1997; 97US-0048964.
 PR 06-JUN-1997; 97US-0048974.
 PR 13-JUN-1997; 97US-0049610.
 PR 08-JUL-1997; 97US-0051926.
 PR 16-JUL-1997; 97US-0052874.
 PR 18-AUG-1997; 97US-0055724.
 PR 22-AUG-1997; 97US-0056630.
 PR 22-AUG-1997; 97US-0056631.
 PR 22-AUG-1997; 97US-0056632.
 PR 22-AUG-1997; 97US-0056636.
 PR 22-AUG-1997; 97US-0056637.
 PR 22-AUG-1997; 97US-0056662.
 PR 22-AUG-1997; 97US-0056664.
 PR 22-AUG-1997; 97US-0056845.
 PR 22-AUG-1997; 97US-0056846.
 PR 22-AUG-1997; 97US-0056847.
 PR 22-AUG-1997; 97US-0056876.
 PR 22-AUG-1997; 97US-0056877.
 PR 22-AUG-1997; 97US-0056878.
 PR 22-AUG-1997; 97US-0056879.
 PR 22-AUG-1997; 97US-0056880.
 PR 22-AUG-1997; 97US-0056881.
 PR 22-AUG-1997; 97US-0056882.
 PR 22-AUG-1997; 97US-0056884.
 PR 22-AUG-1997; 97US-0056886.
 PR 22-AUG-1997; 97US-0056887.
 PR 22-AUG-1997; 97US-0056888.
 PR 22-AUG-1997; 97US-0056889.
 PR 22-AUG-1997; 97US-0056892.
 PR 22-AUG-1997; 97US-0056893.
 PR 22-AUG-1997; 97US-0056894.
 PR 22-AUG-1997; 97US-0056903.
 PR 22-AUG-1997; 97US-0056908.
 PR 22-AUG-1997; 97US-0056909.
 PR 22-AUG-1997; 97US-0056910.
 PR 22-AUG-1997; 97US-0056911.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057659.
 PR 05-SEP-1997; 97US-0057761.
 PR 12-SEP-1997; 97US-0058785.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Bednarik DP, Brewer LA, Carter KC, Duan R, Ehner R, Endress GA,
 PI Feng P, Ferlie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
 PI Kyaw H, Lafleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
 XX
 XX WPI: 1998-506364/43.
 DR P-PSDB: AAM75014.
 XX
 PT New isolated human genes and the secreted polypeptide(s) they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 1: Page 530-531; 721pp; English.
 XX
 CC This sequence represents a nucleic acid molecule designated Gene 171
 CC from the human cDNA clone HTEDJ34 (deposited as clone ATCC 97904 and
 CC ATCC 209050) which encodes a secreted human protein. The gene can be

CC used to generate fusion proteins by linking to the gene to a human
 CC immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of
 CC the fused protein as compared to the human protein only.
 CC The invention relates to 186 novel genes and their fragments (nucleic
 CC acid sequences: AAV59511-59812; amino acid sequences AAM74731-W75026)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 186
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAV59511 for described uses).
 CC
 XX
 XX Sequence 997 BP; 275 A; 213 C; 243 G; 262 T; 4 other;

Query Match 50.0%; Score 942.2; DB 19; Length 997;
 Best Local Similarity 99.6%; Pred. No. 2.1e-293;
 Matches 963; Conservative 2; Mismatches 0; Indels 2; Gaps 2;

OY 869 aggtatgtagagacacattacacccaacaaagaagaagatccctccataatt 928
 DB 1 aggtatgtagagacacattacacccaacaaagaagaagatccctccataatt 60
 OY 929 taactatgtttacaaggatgctgtacattggtctcccgag-atttgcacaattgt 987
 DB 61 taactatgtttacaaggatgctgtacattggtctcccgagatttgcacaattgt 120
 OY 988 ttgaagaacccctaatacccaacaactgattgaattggttaaaagaacttatagccagat 1047
 DB 121 ttgaagaacccctaatacccaacaactgattgaattggttaaaagaacttatagccagat 180
 OY 1048 gaacaccccttgagccaccccttcagatgctgacggtgagtcgctgctgtcccaaac 1107
 DB 181 gaacaccccttgagccaccccttcagatgctgacggtgagtcgctgctgtcccaaac 240
 OY 1108 cccaagtagacatc-tacagatgactctatctccagagctgtgcaagtggcaggtgta 1166
 DB 241 cccaagtagacatccttcagatgactctatctccagagctgtgcaagtggcaggtgta 300
 OY 1167 tgaaggagacatcgaatgaaggtgtcctttagctcccgctctgtgaaatccacagcggc 1226
 DB 301 tgaaggagacatcgaatgaaggtgtcctttagctcccgctctgtgaaatccacagcggc 360
 OY 1227 tatctgcctttaggggcttgaggacttgaattgagttgcttcaaaacatccacttctg 1286
 DB 361 tatctgcctttaggggcttgaggacttgaattgagttgcttcaaaacatccacttctg 420
 OY 1287 caacaagttagcccaagaagttagatgaatgctcttcaagttgcttagaagaatactaac 1346
 DB 421 caacaagttagcccaagaagttagatgaatgctcttcaagttgcttagaagaatactaac 480
 OY 1347 ttataagggcatctatggaacttgaactttgagacacactatgagagcgttctactctgt 1406
 DB 481 ttataagggcatctatggaacttgaactttgagacacactatgagagcgttctactctgt 540
 OY 1407 gggcagaagcagtgatacaaatgctcagaacttgccttggaacagtttggttggaacag 1466
 DB 541 gggcagaagcagtgatacaaatgctcagaacttgccttggaacagtttggttggaacag 600
 OY 1467 ggccttgcaatctgtagcatcctttagagataagaaggtgtctatatagcttvggtaagt 1526
 DB 601 ggccttgcaatctgtagcatcctttagagataagaaggtgtctatatagcttvggtaagt 660
 OY 1527 agatctttgctctgcaaatgtgctgctgggtgtaagtgtctgttctcaccctaac 1586
 DB 661 agatctttgctctgcaaatgtgctgctgggtgtaagtgtctgttctcaccctaac 720
 OY 1587 cctaagtagtctcctaacttctcctaagtagaagatgaagaaactctgtatagga 1646
 DB 721 cctaagtagtctcctaacttctcctaagtagaagatgaagaaactctgtatagga 780
 OY 1647 gagtgaagagagatatgtgtgtagagcaattgacttgcattgaaatgcctgtgtgactt 1706

|||||
Db 781 gagtgaaggaggatctgtgtgtagactgatttcaagtgaatgcctctgttagctt 840
OY 1707 ttccattctgtgagctgccttccctaataatccagtttgtagctgtgaagaagact 1766
|||||
Db 841 ttccattctgtgagctgccttccctaataatccagtttgtagctgtgaagaagact 900
OY 1767 ttgattggaagaagaccttccctctgtactgttaacttaataataatagctcctgatt 1826
|||||
Db 901 ttgattggaagaagaccttccctctgtactgttaacttaataataatagctcctgatt 960
OY 1827 caaagta 1833
|||||
Db 961 caaagta 967
RESULT 9
AAV59681
ID AAV59681 standard; DNA; 777 BP.
XX
AC AAV59681;
XX
DT 19-JAN-1999 (first entry)
XX
DE Human secreted protein gene 171 clone HTEDJ34.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; ischaemia; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; retinosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN MO9839448-r2.
XX
PD 11-SEP-1998.
XX
PF 06-MAR-1998; 98MO-US04493.
XX
PR 02-OCT-1997; 97US-0061060.
PR 07-MAR-1997; 97US-0038621.
PR 07-MAR-1997; 97US-0040161.
PR 07-MAR-1997; 97US-0040162.
PR 07-MAR-1997; 97US-0040163.
PR 07-MAR-1997; 97US-0040333.
PR 07-MAR-1997; 97US-0040334.
PR 07-MAR-1997; 97US-0040336.
PR 07-MAR-1997; 97US-0040626.
PR 11-APR-1997; 97US-0043311.
PR 11-APR-1997; 97US-0043312.
PR 11-APR-1997; 97US-0043313.
PR 11-APR-1997; 97US-0043314.
PR 11-APR-1997; 97US-0043358.
PR 11-APR-1997; 97US-0043569.
PR 11-APR-1997; 97US-0043576.
PR 11-APR-1997; 97US-0043578.
PR 11-APR-1997; 97US-0043580.
PR 11-APR-1997; 97US-0043669.
PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0047492.
PR 23-MAY-1997; 97US-0047500.
PR 23-MAY-1997; 97US-0047501.
PR 23-MAY-1997; 97US-0047502.
PR 23-MAY-1997; 97US-0047503.
PR 23-MAY-1997; 97US-0047581.
PR 23-MAY-1997; 97US-0047582.

PR 23-MAY-1997; 97US-0047583.
PR 23-MAY-1997; 97US-0047584.
PR 23-MAY-1997; 97US-0047585.
PR 23-MAY-1997; 97US-0047586.
PR 23-MAY-1997; 97US-0047587.
PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0053724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057669.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;

XX Sequence 524 BP; 171 A; 90 C; 127 G; 136 T; 0 other:
SQ

Query Match 24.2%; Score 456; DB 21; Length 524;
Best Local Similarity 100.0%; Pred. No. 1.8e-136;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1378 gacacactatgagagcgttctactctgtggtggaagagacatgtacaacatgtccagac 1437
DB 1 gacacactatgagagcgttctactctgtggtggaagagacatgtacaacatgtccagac 60
QY 1438 ttgctggtggaagagctgtggtggtggaagagcgttctgcaatctgtgcactcttaggata 1497
DB 61 ttgctggtggaagagctgtggtggtggaagagcgttctgcaatctgtgcactcttaggata 120
QY 1498 agagggctgtctatagatttgggtgaagtgaatcttctgctgcaaatgtgcctggg 1557
DB 121 agagggctgtctatagatttgggtgaagtgaatcttctgctgcaaatgtgcctggg 180
QY 1558 tgaatgtctgttctctcacccttaaccctagtagttctcactaacttctcacta 1617
DB 181 tgaatgtctgttctctcacccttaaccctagtagttctcactaacttctcacta 240
QY 1618 agtgaagaatgagacgtctgtatagaggagagtgaaggagatatgtgttagagcactt 1677
DB 241 agtgaagaatgagacgtctgtatagaggagagtgaaggagatatgtgttagagcactt 300
QY 1678 gatttcagttgaatgtcgtgtgtgttagctttcattctgtgagagctgcgttccataaa 1737
DB 301 gatttcagttgaatgtcgtgtgtgttagctttcattctgtgagagctgcgttccataaa 360
QY 1738 ttccaggtttgtgtagcgtggaagagaaacttgaatgaaagaaacttccctctgtact 1797
DB 361 ttccaggtttgtgtagcgtggaagagaaacttgaatgaaagaaacttccctctgtact 420
QY 1798 gtttaacttaaaataatagctccctgattcaagta 1833
DB 421 gtttaacttaaaataatagctccctgattcaagta 456

RESULT 11

AA24043
ID AA24043 standard; cDNA; 2204 BP.

XX AA24043;

DT 28-JUN-1999 (first entry)

XX Human core 2Gnt DNA.

XX Screening; treatment; prevention; cardiomyopathy; inhibitor;
KM diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;
KM UDP-GlcNAc:Galbetail-3GalNAc-alphaH; transgenic animal; germ line;
KM beta-1,6-N-acetylglucosaminyl-1-transferase; human; core 2 Gnt; ss.

XX Homo sapiens.

XX OS Location/Qualifiers

FT CDS 244..1530

FT /tag= a

FT /product= "core 2 Gnt"

XX CA2186987-A.

XX 02-APR-1998.

XX 02-OCT-1996; 96CA-2186987.

XX 02-OCT-1996; 96CA-2186987.

XX (MOUN) MOUNT SINAI HOSPITAL CORP.

PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE;

XX WPI; 1998-399608/35.

DR P-PSDB; AAW93943.

PT Screening for substances that prevent or treat cardiomyopathy
PT associated with diabetes and hyperglycaemia - comprises reacting
PT core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor
PT in presence of test substance

XX Disclosure; Fig 9; 35pp; English.

CC This invention describes a method for screening for a substance that
CC may be used to prevent or treat cardiomyopathy associated with diabetes
CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
CC (UDP-GlcNAc:Galbetail-3GalNAc-alphaH-beta-1,6-N-acetylglucosaminyl-
CC transferase) with an acceptor substrate and a sugar nucleotide donor in
CC the presence of a test substance under conditions whereby the core 2
CC GlcNAc-T produces a reaction product, determining the amount of reaction
CC product, and comparing the amount of reaction product with the amount
CC obtained in the absence of the test substance, where lower amounts of
CC reaction product in the presence of the test substance indicate that the
CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)
CC methods for preventing or treating cardiomyopathy associated with
CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
CC germ cells and somatic cells all contain a DNA construct introduced into
CC the animal or an ancestor of the animal at an embryonic stage, where
CC incorporation of the DNA construct into the germ line of the animal
CC causes the animal to develop cardiomyopathy similar to that associated
CC with diabetes mellitus and hyperglycaemia. This sequence encodes the
CC human core 2 Gnt protein (beta-1,6-N-acetylglucosaminyltransferase)
CC which is used to describe the method of the invention.

XX Sequence 2204 BP; 641 A; 414 C; 498 G; 651 T; 0 other:
SQ

Query Match 19.6%; Score 369.4; DB 19; Length 2204;
Best Local Similarity 60.2%; Pred. No. 4.4e-108;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

QY 337 aagaagcgaagagcctttcacagacaccacactactctccctccacagagctgtgagac 396
DB 490 aagaagcgaagagcctttcacagacaccacactactctccctccacagagctgtgagac 549
QY 397 ttcaaggtctgaagagatcatalacagttcccaatgagcaagaaggtgtgtccct 456
DB 550 ttcaaggtctgaagagatcatalacagttcccaatgagcaagaaggtgtgtccct 609
QY 457 atgcatactctatgtgtatcattgtagaagattgaaacttgaagcctgcagact 516
DB 610 atgcatactctatgtgtatcattgtagaagattgaaacttgaagcctgcagact 669
QY 517 gtgtagccctcagaacataactgtgtccatgtgtagatgagaagctccccaagaacttc 576
DB 670 atctatagctccagaacattctatgtgtccatgtgtagatgagaagctccccaagaacttc 729
QY 577 aaagagcgtgcaaaagcaattatttctgtccccaatgtcttccatagccagtaagctg 636
DB 730 tttagcgcagtgatgtgacatgcgtctctgttagtaagtcttctgtgcccagcagatg 789
QY 637 gtccggtgtgttattcctctctgtccaggtgtgcaagctgacccaactgtgtagaagac 696
DB 790 gtccggtgtgttattcctctctgtccaggtgtgcaagctgacccaactgtgtagaagac 849
QY 697 ttgtctcagagctcagtgccgtgtgaaatacttctcgaatacagtgtggcagcgttct 756
DB 850 ctctatgcaatgtgagtgcaactgtgaaatctgtgtaatacttctgtgtgtgttcc 909
QY 757 ataaagcaatgagagatgtgtccaggtcctcaagatgtgtgtagtggagaaatagatg 816
DB 910 ataaagcaatgagagatgtgtccaggtcctcaagatgtgtgtagtggagaaatagatg 969
QY 817 gagtccaggtgtactcctcctaagcaaaagaaacccgctggaatatatacttggagtgtg 876

Db	970	gaacacgagagagatgccaatccccaataagagaagatggaagaagcgtatagatgcgt	10232
Qy	877	agagacacatathacaaccaacaagaagaagatctcccccattatattaaactag	936
Db	1030	aatggaagcgtgacaacacacagagactgtccaatgcttccccaactcgaacacactc	10859
Qy	937	tttcacgaggaatcgtctacatctgtgcttcccgagatctgcaccaatggtttgaagaac	996
Db	1090	tttctctgcagttgcctactctcgtctgctgacgtaggtatggttggttactacaagat	1149
Qy	997	ccataatcccaacacacgtatggaatggttaagaagaacttatagaaccagatgaacactc	1056
Db	1150	gaaaaatcatcaaaaagtgtgagaggtggcacaagaacacataagccctgatagtatctc	1209
Qy	1057	tggggccaccccttcacgcgtgacaggtgagatgctgtgctgtgttcccaaccacccacatgac	1116
Db	1210	tgggcacacatccaagaagattcctgaagtcctcggtgcactacccctgcacgacataagat	1269
Qy	1117	gacatctcagacacagactcttatctatctgcacagcgtgtcaagtgtgcagaggtctcatagtagagagac	1176
Db	1270	gactatctcagacacagcagaagatgtgcacaggtttgtcaaatgtgcatacttttgaagtgat	1328
Qy	1177	atcgataaggggtgctctcttatgtccctcgtgtctcgaaatccacacagcgggtatctgcgtt	1236
Db	1330	gttcccaaggggtgctccctccctcccgccctcgatagtgatgccatggtgcctcagtgatc	1388
Qy	1237	tatggagcctgggagacttgaattgagatgctcaaaaacacatcacactgttgcgacaacagtt	1296
Db	1390	ttcggagctggtgactgttgactgagatgctgacgtgcgaaacaccactgtgttgcacaatagtt	1449
Qy	1237	gacccaaggtagatgataatgctctcactcagtgcttgagaagaatacctaactatataagc	1355
Db	1450	gacgttgatgttctgaccctcttcgacacatggtttgagatgacatcttgagacaacaaag	1508

PR	02-OCT-1996;	96US-0046876.	PA (KING/) KING G L.	PI Nishio Y, Koya D, King GL, Warren CE, Dennis JW;
PN	US6131578-A.		PA (NISH/) NISHIO Y.	
PD	17-OCT-2000.		PA (KOYA/) KOYA D.	
PF	02-OCT-1997;	97US-0943058.	PA (DENN/) DENNIS J W.	
PG			PA (WARR/) WARREN C E.	
PH				
PI				
PJ				
PK				
PL				
PM				
PN				
PO				
PP				
PQ				
PR				
PS				
PT				
PV				
PW				
PX				
PY				
PZ				
QA				
QB				
QC				
QD				
QE				
QF				
QG				
QH				
QI				
QJ				
QK				
QL				
QM				
QN				
QO				
QP				
QQ				
QR				
QS				
QT				
QU				
QV				
QW				
QX				
QY				
QZ				
RA				
RB				
RC				
RD				
RE				
RF				
RG				
RH				
RI				
RJ				
RK				
RL				
RM				
RN				
RO				
RP				
RQ				
RS				
RT				
RU				
RV				
RW				
RX				
RY				
RZ				
SA				
SB				
SC				
SD				
SE				
SF				
SG				
SH				
SI				
SJ				
SK				
SL				
SM				
SN				
SO				
SP				
SQ				
SR				
SS				
ST				
SU				
SV				
SW				
SX				
SY				
SZ				
TA				
TB				
TC				
TD				
TE				
TF				
TG				
TH				
TI				
TJ				
TK				
TL				
TM				
TN				
TO				
TP				

[illegible]

Db 1426 gacgtgagatgttgcaccttgcacatccagtggttgatgagacatttgagacacaaagc 1484

RESULT 14

AAAX24042

AAAX24042 standard; cDNA: 5010 BP.

AAAX24042;

28-JUN-1999 (first entry)

Rat DH1 cDNA.

DH1; rat; screening; treatment; prevention; cardiomyopathy; inhibitor; diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate; UDP-GlcNAc:galbetail-3galNAc-alphaHAR; transgenic animal; germ line; beta-1,6-N-acetylglucosaminyl-transferase; SS.

Rattus sp.

Location/Qualifiers

Key 802..2088

/tag= a

/product= "DH1"

CA2186987-A.

02-APR-1998.

02-OCT-1996; 96CA-2186987.

02-OCT-1996; 96CA-2186987.

(MOUNT) MOUNT SINAI HOSPITAL CORP.

Dennis JW, King GL, Koya D, Nishio Y, Warren CE;

WP1; 1998-399608/35.

P-PSDB; AAM93942.

Screening for substances that prevent or treat cardiomyopathy associated with diabetes and hyperglycaemia - comprises reacting core 2 GlcNAc-T in presence of test substance

Example 1; Fig 3A; 35pp; English.

This invention describes a method for screening for a substance that may be used to prevent or treat cardiomyopathy associated with diabetes and hyperglycaemia. This method involves reacting core 2 GlcNAc-T (UDP-GlcNAc:galbetail-3galNAc-alpha beta-1,6-N-acetylglucosaminyl-transferase) with an acceptor substrate and a sugar nucleotide donor in the presence of a test substance under conditions whereby the core 2 GlcNAc-T produces a reaction product, determining the amount of reaction product, and comparing the amount of reaction product with the amount obtained in the absence of the test substance, where lower amounts of reaction product in the presence of the test substance indicate that the substance inhibits core 2 GlcNAc-T. The invention also describes (1) methods for preventing or treating cardiomyopathy associated with diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose germ cells and somatic cells all contain a DNA construct introduced into the animal or an ancestor of the animal at an embryonic stage, where incorporation of the DNA construct into the germ line of the animal causes the animal to develop cardiomyopathy similar to that associated with diabetes mellitus and hyperglycaemia. This sequence encodes the rat DH1 protein which is used to describe the method of the invention.

Sequence 5010 BP; 1369 A; 987 C; 1093 G; 1561 T; 0 other;

Query Match

Best Local Similarity 58.3%; Score 364.6; DB 19; Length 5010; Matches 659; Conservative 0; Mismatches 469; Indels 3; Gaps 1;

QY 251 caaagaggtctatcaactgttcaagggtccacgaggggagccaaagagcagtgctcagg 310
 Db 959 ccaataagcaatgttattatgaccacaaagtcttacagggtgacccagagaatccagaagg 1018
 QY 311 ctatcttgataaactgtgaagtcac---gaagaagcagagccttccacagaccact 367
 Db 1019 tgaagcttgagataactaaacagtcgaatcaagaagcgtccgagcgagacccagatgc 1078
 QY 368 acctctccctacacagagagcttgagcacttcaaggctgaaaggagatcatcagcttcc 427
 Db 1079 atataaactgaccccgagctgcctccctcatcagaacagcaaatatattatgagc 1138
 QY 428 cactgagcaagaagagtgaggttccctatcactctatctgttctcatttgatgaaaga 487
 Db 1139 cctctactcaaaagaagagtggtgttcttccatgtcatctatctatgtgtctcactaaga 1198
 QY 488 ttgaaacttgaaaggtcactgtcgagctgtgtatgcccctcagaacatactgtgtcc 547
 Db 1199 ttgacatgttgacagcagctcctgagggccatctatatagcaagaatttctactgtatc 1258
 QY 548 atgttgatgagagagtcgcccaagaacttcaagaagcggtcaagaatattatttctgt 607
 Db 1259 acgttgacagaaagcagaggaalcccttctttagccgcggtgacaggtcatcctcgtc 1318
 QY 608 tcccaatgtctcctatagcagtaagctgttcggtgtgttattgtccctcgtccagag 667
 Db 1319 ttgataaagtcttcttggtgcagccaggttgagagtggtgtataagcctcctcgtgagtc 1378
 QY 668 tgcagagctgacactcaactgacatgagagactgtcccaagtcagtgccgtgagaaact 727
 Db 1379 ttaagctgacactcaactgacatgagagactgtacagaaatgataactgagagtaact 1438
 QY 728 tcttgaatacagtggtgagcagacttccctataaagacaaatgacagatggtccagctc 787
 Db 1439 tgaatcaatcttggtgtatgagattccctataaaccacactgaaatgttcaagaagc 1498
 QY 788 tcaagatgttgatgaggaataagatgagtgagtgagctacccctaaagcaagaaga 847
 Db 1499 tcaagctcctcagagggagaaacagcctggaacatgagagagtgctcccaacagaaga 1558
 QY 848 ccgcgtgaaatataactttgagtagtgaagacacattacacctaacaagaaga 907
 Db 1559 agagtggtgaaacacgatacacggtgtgagcaggaagtgacaaacatgagtgatca 1618
 QY 908 aggatccctccctataataattactatgtttaacagggagatgctgacatgtgtgtccc 967
 Db 1619 aggcgcagctccactcaaaactcctctcttccagcagcgccatttcgtgtcacta 1678
 QY 968 gagatttcgtccaacatgttttgaagaaccctaaatcccaacactgattgaatgata 1027
 Db 1679 gggatatagtatgactatgtcgtggaataaataatcaataaagtatcagatgagtgac 1738
 QY 1028 aagaactatagccacagatgaacacctctggtgcaaccctcagcgtgacggtgagatgc 1087
 Db 1739 agaacacatacagcccgagtggtctcctgtggccacacacacacagagatccctgaagtc 1798
 QY 1088 ctggtctgttcccaacaccccaagtcagacatctcagactgtacttattgtccagagc 1147
 Db 1799 ctggttctcctccctcaagcacaatagatgactgtctgacatgaaatgtctgctgaagt 1858
 QY 1148 tggtaagtgagaggtgtacatgagagacatcgaataaggtctcctctgtccctgtgt 1207
 Db 1859 ttgtcaagtggacaataacttgagagagcagatgttccaaatggcgcgtctatccacgtgca 1918
 QY 1208 ctggaatccacacagcgggtactcgtttatgggctggtggagacttgatgtatgtctc 1267
 Db 1919 gtgagtcacatgtgctgt 1978
 QY 1268 aaacacatcactgtgtgccaacaggtttgaccacaaggtatgataatgtcttgcgt 1327
 Db 1979 gcaaacacacacttlttcccaataagttgacatgagatgtgatactccttgcctccag 2038

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 26, 2002, 04:25:28 ; Search time 94.91 Seconds
(without alignments)
4875.920 Million cell updates/sec

Title: US-09-874-390-1_COPY_436_2319
Perfect score: 1884
Sequence: 1 gatgtgtccctccaccct.....ataataatcaacaga 1884

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_MN:*
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1798.8	95.5	2128	3	US-09-233-506-1 Sequence 1, Appl1
2	369.4	19.6	2105	1	US-07-955-041-3 Sequence 3, Appl1
3	369.4	19.6	2105	1	US-08-227-455-3 Sequence 3, Appl1
4	369.4	19.6	2105	1	US-08-472-482-3 Sequence 3, Appl1
5	369.4	19.6	2105	1	US-08-487-069-3 Sequence 3, Appl1
6	355	18.8	2102	3	US-09-063-237-3 Sequence 3, Appl1
7	180.4	9.6	1807	1	US-08-118-906-13 Sequence 13, Appl1
8	180.4	9.6	1807	1	US-08-486-196-13 Sequence 13, Appl1
9	180.4	9.6	1807	1	US-08-488-135-13 Sequence 13, Appl1
10	180.4	9.6	1807	2	US-08-474-065-13 Sequence 13, Appl1
11	162.8	8.6	378	1	US-08-118-906-3 Sequence 3, Appl1
12	162.8	8.6	378	1	US-08-486-196-3 Sequence 3, Appl1
13	162.8	8.6	378	1	US-08-488-135-3 Sequence 3, Appl1
14	162.8	8.6	378	2	US-08-474-065-3 Sequence 3, Appl1
15	134	7.1	378	1	US-08-118-906-1 Sequence 1, Appl1
16	134	7.1	378	1	US-08-486-196-1 Sequence 1, Appl1
17	134	7.1	378	1	US-08-488-135-1 Sequence 1, Appl1
18	134	7.1	378	2	US-08-474-065-1 Sequence 1, Appl1
19	129.6	6.9	192	3	US-09-233-506-9 Sequence 9, Appl1
20	84	4.5	147	3	US-09-233-506-13 Sequence 13, Appl1
21	43.6	2.3	66	1	US-08-118-906-11 Sequence 11, Appl1
22	43.6	2.3	66	1	US-08-486-196-11 Sequence 11, Appl1
23	43.6	2.3	66	1	US-08-488-135-11 Sequence 11, Appl1
24	43.6	2.3	66	2	US-08-474-065-11 Sequence 11, Appl1
25	43	2.3	99	1	US-08-118-906-7 Sequence 7, Appl1
26	43	2.3	99	1	US-08-486-196-7 Sequence 7, Appl1
27	43	2.3	99	1	US-08-488-135-7 Sequence 7, Appl1

28	43	2.3	99	2	US-08-474-065-7 Sequence 7, Appl1
29	35	1.9	99	1	US-08-118-906-5 Sequence 5, Appl1
30	35	1.9	99	1	US-08-486-196-5 Sequence 5, Appl1
31	35	1.9	99	1	US-08-488-135-5 Sequence 5, Appl1
32	35	1.9	99	2	US-08-474-065-5 Sequence 5, Appl1
33	33.2	1.8	1941	4	US-09-402-002-1 Sequence 1, Appl1
34	33	1.8	11283	3	US-08-603-753D-3 Sequence 3, Appl1
35	33	1.8	11283	3	US-09-099-753-3 Sequence 3, Appl1
36	33	1.8	11283	4	US-08-986-106-3 Sequence 3, Appl1
37	32.8	1.7	246240	2	US-08-724-394A-20 Sequence 20, Appl1
38	32.8	1.7	246240	2	US-08-724-394A-21 Sequence 21, Appl1
39	32.8	1.7	246240	2	US-08-724-394A-22 Sequence 22, Appl1
40	32.2	1.7	3164	1	US-07-723-002C-3 Sequence 3, Appl1
41	32.2	1.7	7218	1	US-08-232-463-14 Sequence 14, Appl1
42	31.8	1.7	900	1	US-08-218-026-1 Sequence 1, Appl1
43	31.8	1.7	900	2	US-08-653-632-1 Sequence 1, Appl1
44	31.8	1.7	1653	1	US-08-218-026-3 Sequence 3, Appl1
45	31.8	1.7	1653	2	US-08-653-632-3 Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-09-233-506-1
; Sequence 1, Application US/09233506
; Patent No. 6136580
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Yeh, Jiumn-Chern
; TITLE OF INVENTION: A Beta1-1-6-N-Acetylglucosaminyltransferase That Forms
; TITLE OF INVENTION: Core 2, Core 4 and I Branches
; FILE REFERENCE: P-1J 3415
; CURRENT APPLICATION NUMBER: US/09/233, 506
; CURRENT FILING DATE: 1999-01-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2128
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (354)..(1670)
US-09-233-506-1

Query Match 95.5%; Score 1798.8; DB 3; Length 2128;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1812; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY	1	gatgtgtccctccaccctccctctgtgtcgtcgtccaccgttcccatctgtgagc	60
DB	294	gatgtgtccctccctccaccctccctctgtgtcgtcgtccaccgttcccatctgtgagc	353
QY	61	atgtgtcaatgaagaagagctctgcagctgcattactgtgtgctgtggtctataty	120
DB	354	atgtgtcaatgaagaagagctctgcagctgcattactgtgtgctgtggtctataty	413
QY	121	ctgtgtgaccatgtgtgcttgaacattctcttcaggttggaagtgtgacctgtgac	180
DB	414	ctgtgtgaccatgtgtgcttgaacattctcttcaggttggaagtgtgacctgtgac	473
QY	181	ggtctgtgagtcacgaagaaatctcaagcagctgtaggaatctgtataatttcctg	240
DB	474	ggtctgtgagtcacgaagaaatctcaagcagctgtaggaatctgtataatttcctg	533
QY	241	aaacttcagaagaagagcttatacaactgttcagggtgcaccgagggagcaagagca	300
DB	534	aaacttcagaagaagagcttatacaactgttcagggtgcaccgagggagcaagagca	593
QY	301	gtctgtcagctattctgaataacctgtgaggtcaagaagcgagagccttcacagac	360

Db 594 gtgtctcaagctattctgtaaacctgaggtcaagaagaagcgaagccttcaagac 653
QY 361 acccaataacctctccctaccagaagactgtgagcaacttcaaggctgaagaagttcata 420
Db 654 acccaataacctctccctaccagaagactgtgagcaacttcaaggctgaagaagttcata 713
QY 421 cagttcccaactgagcaagaagaggtgaggttccctattgcatactctatgtgattcat 480
Db 714 cagttcccaactgagcaagaagaggtgaggttccctattgcatactctatgtgattcat 773
QY 481 gagaagaattgaaaactttgaaaagctacttcgagctgtgatatccctccgaacataaac 540
Db 774 gagaagaattgaaaactttgaaaagctacttcgagctgtgatatccctccgaacataaac 833
QY 541 tggatcatgttgatgagaagctcccaagaacttcaaaaagcggtcaagaacattat 600
Db 834 tggatcatgttgatgagaagctcccaagaacttcaaaaagcggtcaagaacattat 893
QY 601 tcttgcttcccaaatgtcttcataagccaagtgttcgggttggtttaaagctccctg 660
Db 894 tcttgcttcccaaatgtcttcataagccaagtgttcgggttggtttaaagctccctg 953
QY 661 tccaaggttgagaactgacctcaactgcatggaagactgtctcagaagctgaagtcgtg 720
Db 954 tccaaggttgagaactgacctcaactgcatggaagactgtctcagaagctgaagtcgtg 1013
QY 721 aactactctcgtatacatgttggaacgacttccctaataagaagcaatgcaagatgtc 780
Db 1014 aactactctcgtatacatgttggaacgacttccctaataagaagcaatgcaagatgtc 1073
QY 781 caggtctcacaagatgttgaatggagggaatagcatggaagtcagaagtaacctcctaagac 840
Db 1074 caggtctcacaagatgttgaatggagggaatagcatggaagtcagaagtaacctcctaagac 1133
QY 841 aaaaaaacccgcgggaataatctactttagttagtgaagacacataacctcaaaccaac 900
Db 1134 aaaaaaacccgcgggaataatctactttagttagtgaagacacataacctcaaaccaac 1193
QY 901 aaaaaaagaagatctccccccttaataattactatgttttcaaggaatgcatcatgtg 960
Db 1194 aaaaaaagaagatctccccccttaataattactatgttttcaaggaatgcatcatgtg 1253
QY 961 gcttcccgagattcgtccaacatgttttgaagaacccctaatacccaacatgattgaa 1020
Db 1254 gcttcccgagattcgtccaacatgttttgaagaacccctaatacccaacatgattgaa 1313
QY 1021 tgggttaaaaagacattatagccagatgaaacctctgggccaaccttcaagtcgacgg 1080
Db 1314 tgggttaaaaagacattatagccagatgaaacctctgggccaaccttcaagtcgacgg 1373
QY 1081 tggatgctgtgctctgttcccaacaccccaagtaagacatctcagaacatgattctat 1140
Db 1374 tggatgctgtgctctgttcccaacaccccaagtaagacatctcagaacatgattctat 1433
QY 1141 gccaggtgtgtaagtggcagggtcatgaaaggagacatcgataagggtgctccttatgtc 1200
Db 1434 gccaggtgtgtaagtggcagggtcatgaaaggagacatcgataagggtgctccttatgtc 1493
QY 1201 cctgtcttggaatccaacggcggtatctggtttatgggctggggagctggaattgg 1260
Db 1494 cctgtcttggaatccaacggcggtatctggtttatgggctggggagctggaattgg 1553
QY 1261 atgcttcaaaaacatacactgttgcgaacaagtttgacccaagaagttagatgataatgtc 1320
Db 1554 atgcttcaaaaacatacactgttgcgaacaagtttgacccaagaagttagatgataatgtc 1613
QY 1321 ctccaagtgtctagaagaataactaagttataaggccaatctatgaggaacttgagac 1380
Db 1614 ctccaagtgtctagaagaataactaagttataaggccaatctatgaggaacttgagac 1673
QY 1381 acaactatgaagcgttgcactcgttggggaagaagcatgtaaaaaatgctcgaacttg 1440
Db 1674 acaactatgaagcgttgcactcgttggggaagaagcatgtaaaaaatgctcgaacttg 1733

QY 1441 ctgggaacgtgtgtgtgtggagaccaggcttgcnaattcgtggcatcctttaagataaga 1500
Db 1734 ctgggaacgtgtgtgtgtggagaccaggcttgcnaattcgtggcatcctttaagataaga 1793
QY 1501 gggctgtatataatgtgtgtgtgaatagatcctttgcttgcgaatattgctgctgggtga 1560
Db 1794 gggctgtatataatgtgtgtgtgaatagatcctttgcttgcgaatattgctgctgggtga 1853
QY 1561 atgctgtgtctctcaacccctcaacccctagtagttccctccataacttctcaactgaat 1620
Db 1854 atgctgtgtctctcaacccctcaacccctagtagttccctccataacttctcaactgaat 1913
QY 1621 gagaatgagaacgtctgtgaataaggagatgaaaggaggaatagatgtgtgaagcaactgtat 1680
Db 1914 gagaatgagaacgtctgtgaataaggagatgaaaggaggaatagatgtgtgaagcaactgtat 1973
QY 1681 ttcaagtgaatgcctgtcgtgtgaactttccatctgttgaagctgcggttccataatc 1740
Db 1974 ttcaagtgaatgcctgtcgtgtgaactttccatctgttgaagctgcggttccataatc 2033
QY 1741 caggtttgtagcgttgaggagaactttgtagaagaagacacttccctctgtactgtt 1800
Db 2034 caggtttgtagcgttgaggagaactttgtagaagaagacacttccctctgtactgtt 2091
QY 1801 aacttaaaaataa 1816
Db 2092 aacttaaaaataa 2107

RESULT 2
US-07-955-041-3
Sequence 3, Application US/07955041
Patent No. 5360733

GENERAL INFORMATION:

APPLICANT: FUKUDA, MINORU
APPLICANT: BIERHUIZEN, MARTI FA
TITLE OF INVENTION: A NOVEL BETAL-6
TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
TITLE OF INVENTION: LEUCOSTALIN AND A METHOD FOR CLONING PROTEINS HAVING
TITLE OF INVENTION: ENZYMATIC ACTIVITY

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESSES:

ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
CITY: SAN DIEGO

STATE: CALIFORNIA
COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/955,041

FILING DATE: 19921001

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: CAMPBELL, CATHRYN
REGISTRATION NUMBER: 31,815
TELECOMMUNICATION NUMBER: P-LJ 9294

TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949

INFORMATION FOR SEQ. ID NO. 3:
SEQUENCE CHARACTERISTICS:

LENGTH: 2105 base pairs
TYPE: NUCLEIC ACID

STRANDEDNESS: both
TOPOLOGY: linear

MOLECULE TYPE: cDNA
FEATURE:


```

: NAME/KEY: CDS
: LOCATION: 220..1504
: FEATURE:
: NAME/KEY: polyA_signal
: LOCATION: 1913..1918
: FEATURE:
: NAME/KEY: misc_signal
: LOCATION: 248..314
: OTHER INFORMATION: /standard_name=
: OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-07-955-041-3

```

```

Query Match          19.6%: Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%: Pred. No. 5.9e-111;
Matches 613: Conservative 0; Mismatches 406; Indels 0; Gaps 0;

```

```

Oy 337 aagaagcgaagccttcacagaccccaactcctccctcacaagaagctgtgagac 396
Db 466 AAAAAAGCCCGTCGTGACACCTGACACTATATATAACAGACAGAGCTCTCTCT 525
Oy 397 ttcaagctgaagaagctatatacagttccctcagcagaagaagctgtgagct 456
Db 526 TTCAATCAAGAGACGCAATATATTTGTAGAACCCCTTACTAAGAAGAGCGAGTTTCA 585
Oy 457 atgcatactatgtgtatcatgagaagaattgaaacttgaagaagctcagagct 516
Db 586 ATAGCATATCTTATAGTGTTCATCACAAGATGAAATGCTTGACAGGCTGCTGAGGCG 645
Oy 517 gtgtatgccctcagaacatactatgtgtcatgtgagtgagaagctcccaagaacttc 576
Db 646 ATCTATATGCTCAGAAATTTCTATTTGCTTCATGTGACACAAATCCGAGGATTCCTAT 705
Oy 577 aagaagcgcgcgaagaagaactatcttctcctcccaaatgtctcatagcagaagtaagctg 636
Db 706 TTAGCTGAGTGAATGGGATGCTTCTCTTTTACTATATGCTTTGTGTGCGACCGGANTG 765
Oy 637 gtctgggtgttattgtccctcgtgtccaggtgtcgaagctcctcaactgtatgagaac 696
Db 766 GAGAGTGTGTTTATGATCGTGGAGCGGGTTCAAGCTGACCTCAATGCATGAAGAT 825
Oy 697 ttgtctcagagctcaagtcggtggaataacttctcgtgaatatactgtggaagcgaacttc 756
Db 826 CTCTATGCAATGAGTGCACAACTGGAAGTACTTGATTAATCTTTGTGATGATTTTCC 885
Oy 757 ataaagcaatgtcagaatgtgtcaggtctcagaatgtgtgaatgtgaagaatagctg 816
Db 886 ATTAACCAACCTAGAAATTTGTCAAGAGCTCAAGTTGTATATGGAGAAACCAACTG 945
Oy 817 gactcagaggtactcctcctaagcacaagaacccgtgtgaataatacacttgaagtg 876
Db 946 GAAACGAGAGGATGTCCTCCATTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1005
Oy 877 agagacacattacacctaacaagaagaagatcctcccttataactaactatg 936
Db 1006 AATGGAAGCTGACAAACACAGGAGCTGTCAAAATGCTTCCCTCCACTGAAACACCTCTC 1065
Oy 937 tttaaggaatgtgtatctgtgtgtcctccgagattcgtccaaatglttgaagaac 996
Db 1066 TTTCTGGACATGCTTACTTGTGTGTCAGTAGGAGTATGTGGGTATGCTACTACGAAT 1125
Oy 997 cctaactccacaactgtatgaaatgtgaaagaacttlaagccagaatgaacaactc 1056
Db 1126 GAAAAAATTCAAAAGTTGATGAGTGGGACACAAAGACATACAGCCCTGATGAGATCTC 1185
Oy 1057 tgggcaacctcagcgtgcaagtgagtgagctgtgtcgttcccaaccacccaaglac 1116
Db 1186 TGGGCAACATCAAAAGATTCCTGAAGTCCGGGCTCAGTCCCGCAGCAGCATTAAGTAT 1245
Oy 1117 gaactctagaagactactatgtccagctgtgtcgaagtgccaggtgtcaggaagaac 1176
Db 1246 GATCTATCTGACATGCAAGAGGTGCGAGTTTGTCAAGTGGCAGTACTTGTGAGGAT 1305

```

```

Oy 1177 atcgaagggtgtcctcctatgtcctccctcgtcgtgaatccacacagcggtatctgcgt 1236
Db 1306 GTTCCAAAGGGTGTCCCTACCGCGCATGAGTGCATGTGCTGAGTGAAT 1365
Oy 1237 tatgggctgggagcttgaatgtgattcctcaaaacacactgtgtggcaacaagttc 1296
Db 1366 TTCGAGCTGTGACTTGAATGATGATGCTGGCAGAACACCACTTGTGGCAATTAAGTTT 1425
Oy 1297 gaccacaagtgatgataaagctcttcagtgcttagaagaatactcgttataggc 1355
Db 1426 GACGTGATGTGACCTCTTGCATCCAGTGTGTGATGACATTTGAGACACCAAGC 1484

```

RESULT 3

```

US-08-227-455-3
Sequence 3, Application US/08227455
Patent No. 5624832

```

GENERAL INFORMATION:

```

APPLICANT: FUKUDA, MINORU
APPLICANT: BIERHUIZEN, MARTI FA
TITLE OF INVENTION: A NOVEL BETA1-6
TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122

```

COMPUTER READABLE FORM:

```

MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/227,455
FILING DATE: 14-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATHERYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9957
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949

```

INFORMATION FOR SEQ ID NO: 3:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 2105 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 220..1504

```

FEATURE:

```

NAME/KEY: polyA_signal
LOCATION: 1913..1918
FEATURE:
NAME/KEY: misc_signal
LOCATION: 248..314
OTHER INFORMATION: /standard_name=
OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-08-227-455-3

```

```

Query Match          19.6%: Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%: Pred. No. 5.9e-111;
Matches 613: Conservative 0; Mismatches 406; Indels 0; Gaps 0;

```



```

Db 646 ATCTATATGCTCAGAAATTTCTATTGCTTCATGTGACACAAATCCGAGATTCTTAT 705
Oy 577 aaagagcggtcaagaacatlaattcttcttcccaaatgcttcatagacagtaagctg 636
Db 706 TTAGCTGACGATATGGGATCGCTTCCGTGTTTACTAATGTTTGTGGCCAGCGGATTG 765
Oy 637 gtccgggtgttattatgcttccctggtccaggtgtaacagctgaactcaactgataagac 696
Db 766 GAGAGTGTGTTTATGATCGTGAGCGGGTTGAGGCTGACCTCAACTGATGAAGAT 825
Oy 697 ttgtcccaagagctcagtgccggtggaataacttccgaataactgtgggaagacttccct 756
Db 826 CTCTATGCAATGATGTCAAACTGGAAGTACTGTAATCTTTGTGATGATTTTCC 885
Oy 757 ataagaagaatgcagagatggtccagagctcgaagatgtgaatggagagatagcatg 816
Db 886 ATTAACCAACCTGAAATTTGTCAAGAACTCAAGTTTATATGGAGAAACCACTG 945
Oy 817 gagtcagaggtactcctcctaagacaaagaacccgctggaatatacttggagtagtg 876
Db 946 GAAACGAGAGAGATGCCATCCCATTAAGAAAGATGGAAGACGGTATGAGTGTGTT 1005
Oy 877 agagacacattacactcaacaaagaagaagatctcccttataactatg 936
Db 1006 AATGGAAGCTGCACAAACAGAGGACTGTCAAAATGCTTCCCTCCACTGCAACACTCTC 1065
Oy 937 ttacagaggaatgctgtaactgtgtcccgagatllcccaacatgttttgaagaac 996
Db 1066 TTTTCTGCGACGTGCTTACTTCTGTGTCAGTATGAGGATATGTGGGTATGTACTACAGAT 1125
Oy 997 cctaaccacaacactgaltgaatgggtaaaagaacacttaagccagatgaacaactc 1056
Db 1126 GAAAAAATCAAAAGTTGATGAGTGGGCAACAAGACATACAGCCGTGATAGATATC 1185
Oy 1057 tgggcccacccctcagcggtgacggtgagatgctgtctgttcccaaccccccaagtaac 1116
Db 1186 TGGGCGACCATCAAAAGATTTCTGTAAGTCCGGGCTCACTCCCTCCGACCATTAAGTAT 1245
Oy 1117 gacatctcagacatctatctatgcccagagctggtcgaagtgacaggtgataaggaagac 1176
Db 1246 GATCTATCTGACATCAAGCAATTTGCCAGTTTGTCAAGTGCATCTTATGAGGTGAT 1305
Oy 1177 atcgaataaggtgtcctctatgcttccctgcttgaatcccaacagcggtctatcgtgt 1236
Db 1306 GTTTCAGAGGTGTCCTTACCCGCTGCGATGAGTGCATGTGCGCTCACTGATGCAAT 1365
Oy 1237 tatggagctggagacttgaattgagatgcttcaaacacatcactgttggccaagaattt 1296
Db 1366 TTCCGAGCTGTGACTTAAGTGAAGTGTGCGCAACACCACTTGTTCCTCAATTAAGTTT 1425
Oy 1297 gaccacaaggtagatgaataatgcttcaagtgcttagaagaatatactatagctataagc 1355
Db 1426 GAGCTGATGTTGACCTCTTTGCCATTCAGTGTGTTGATGAGCAATTTGAGACAAAGC 1484

```

US-08-487-069-3
Sequence 3, Application US/08487069

Patent No. 5684134
GENERAL INFORMATION:
APPLICANT: FUKUDA, MINORU
APPLICANT: BIERHUIZEN, MARTI FA
TITLE OF INVENTION: A NOVEL, BETAI-6
TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
TITLE OF INVENTION: ENZYMACTIC ACTIVITY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA

```

; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,069
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/955,041
; FILING DATE: 01-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHERYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9294
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: Linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 220..1504
; FEATURE:
; NAME/KEY: POLYA_signal
; LOCATION: 1913..1918
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 248..314
; OTHER INFORMATION: /standard_name=
; OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
; US-08-487-069-3

```

Query Match 19.6%; Score 369.4; DB 1; Length 2105;
Best Local Similarity 60.2%; Pred. No. 5.9e-11;
Matches 613; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

```

Oy 337 aaagaagcgagagccttccacagaacccccactctccctcaccagagactgtgaagac 396
Db 466 AAAAAAGCGCCCTCGGTGGACACCTGACGACTATATAAACATGACCAAGTGTCTTCT 525
Oy 397 ttcagagctgaaaggaagttcatagcagttccacactgagaagaagaaggtggatccct 456
Db 526 TTCAATCAAGAGACGCAAAATATTTGTAGAACCCCTTAGTAAGAAAGAGAGCGGATTTCCA 585
Oy 457 attgatactctatgtgtatcatagagaagatlgaaacttgaagaagctactgcagact 516
Db 586 ATAGCATATTTATATAGTGGTTCATCAACAGATTGAATGCTTGACAGGCTGCGTAGAGGAT 645
Oy 517 ggtatgcccctcagaacatataactgtgtccatgtgagatgagaagatcccaaaaacttc 576
Db 646 ATCTATATGCTCAGAAATTTCTATTGCTTCATGTGAGACAAAAATCCAGGATTCCTAT 705
Oy 577 aaagagcggtcaagaacatlaattcttcttcccaaatgcttcatagccagtaagctg 636
Db 706 TTAGCTGACGATATGGGATCGCTTCCGTGTTTACTAATGTTTGTGGCCAGCGGATTG 765
Oy 637 gtccgggtgttattatgcttccctggtccaggtgtaacagctgaactcaactgataagac 696
Db 766 GAGAGTGTGTTTATGATCGTGAGCGGGTTGAGGCTGACCTCAACTGATGAAGAT 825
Oy 697 ttgtcccaagagctcagtgccggtggaataacttccgaataactgtgggaagcagacttcc 756
Db 826 CTCTATGCAATGATGTCAAACTGGAAGTACTGTAATCTTTGTGATGATTTTCC 885

```


Db	1245	GATCTAAGTCATGACGACAGAGTTGGCCAGTTTGTCAATGGCACTACTTTCAGGGTAT	1304
Qy	1177	atcgataaagggtgtcctccctatgctccctggtcctctggaatccacaaagcggtatctggtt	1236
Db	1305	GTTTCCAAAGGGTCTCCCTCCCTACCGCCCTCGATGGAGTGTCAATGCGTCAGTGTGCAT	1364
Qy	1237	tatggagcttgggaacttgatctgagcttcaaaacatcactgttggccaacaagttc	1286
Db	1365	TTGCGAGCTGTGACTGTAAGCTGGATGTCGCCAACAACACTTGTTCGCCAATTAAC	1424
Qy	1297	gaccaaaggtagatgaataatgtctctgaagctctgaagaatcactaagttataaagc	1355
Db	1425	GAGCTGATTTGAACCTTTTGGCATTCCAGTGTGTCATGAGCAATTTTGAGACAAAGC	1483

RESULT 7

```

Sequence 13 Application US/08118906
Patent No. 5484590
GENERAL INFORMATION:
APPLICANT: Fukuda, Minoru
APPLICANT: Bierhuizen, Marti F.A.
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human CDNA Encoding a Member of a
TITLE OF INVENTION: Beta 1,6-N-Acetylglucosaminyltransferase Gene Family
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,906
FILING DATE: 09-SEP-1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1807 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 255..1454
US-08-118-906-13

```

Query Match	9.6%	Score 180.4	DB 1	Length 1807
Best Local Similarly	54.3%	Pred. No. 7.3e-49		
Matches 395	Conservative 0	Mismatches 321	Indels 12	Gaps 1

Qy	Dy	Qy	Dy
388	tgtagacacattcaaggtcgaagaagaattcaacgaftccacacgagcaagaagaagtg	447	
468	TGCAAGGAATCTTGACCCAGCAGGCCTACTACTACACCCCTTTATCTAAGGAAGAACT	527	
448	gaattccctattgcatactctatggtgattcatgagaagaattgaaacttgaagcta	507	
528	GACTTTCCCTTGGGATATATATATGTCATATCCATCATACCTTTGACACCTTTGCAAGGCTC	587	

OY	508	ctcgagcctgctgtatagcccccctcagaacatactctgtcccaatgtagaagaagcccca	567
Db	568	TTTCAGGCGCTATTTCATATGCCCCCAAAATATCTACTGTGTTCACTGTGATGATAAAGCAACA	647
OY	568	gaaaccttcaaaagagcgycgtcacaagcaatlattcttcgtccccaatgctcttcac	627
Db	648	ACTGAATTTTAAAGATGCGGTTAGACGACACTATTAAGCTGCTTCCCAAGCGTTTCTGGGT	707
OY	628	agtaagctggtctcgggtggtttagtgccttcctggttcacaggtgtcaagctgaactcagc	687
Db	708	TCCCAAGATGGAACCCGTTGTCTATGAGGAGATCTCCAGGCTCCAGCGTGCACCTGAACTGC	767
OY	688	atggaagactgctcctcagagctcagtcagtcgcgtgtgaaaatactcctcgaatacatgtgaggagc	747
Db	768	ATCAGAGATCTTTCTTCGCTTCGAGGCTCTCANTGGAAGTACGTTATACACCTGTGGGCA	827
OY	748	gacttccatataaagagcaatgcaagatggtgccagcgtctcgaagatgtgaaatggaag	807
Db	828	GACTTTCCCCCTGAACCAACGAAGAAATAGTTCAGTATCTGAAAGATTTTAAAGSTAA	887
OY	808	aatagcatggaagtcagaggtacctccttaagcacaagaaaaccgcctggaatat-----	861
Db	888	AATATCACCCCCAGGGGTCTCGTCCGCCACCTCATGCAATTTGACAGCACTAAATATTTCCAC	947
OY	862	-----caattgagtgtagtgaagagacacattacacccaacaagaagaagatcct	915
Db	948	CNAGAGCACCTGCGCAAAAGAGCTTTCTTATGTGATAGACACACAGCGTTGAACCGCT	1007
OY	916	cccccttaacttaactatglttcaaggaaatgcgtacatbtggtctcccgagatlc	975
Db	1008	CCCCCCCCTAAATCTCCACATTTACTTTTGGCTCTGGCTATGTGGGCTCTATCAGAAGAGTTT	1067
OY	976	gtcccaacatgttttgaagaaccctaaatcccaacaacagatgaatgtgtaaagaacct	1035
Db	1068	GCCAACTTTGTTCTTCATATGACCCACAGGGCTGTGTAATTTGCTTCCATGTGTCACNAGACACT	1127
OY	1036	tatagcccaatgaagacacctctgggcacccacttcagcgtgtgcaagtgtagtgcgtct	1095
Db	1128	TTCACTGCTGTGATGAGCAATTTCTGGGTGACACTCATATAGATTCACAGGTCTCTCGCTCT	1187
OY	1096	gttcccaa 1103	
Db	1188	ATGCCAAA 1195	

RESULT 8

```

US-08-486-196-13
: Sequence 13, Application US/08486196
: Patent No. 5731420
:
: GENERAL INFORMATION:
:
: APPLICANT: Fukuda, Minoru
:
: APPLICANT: Bierhuizen, Marti F.A.
:
: TITLE OF INVENTION: Expression of the Developmental I
:
: TITLE OF INVENTION: Antigen By a Cloned Human CDNA Encoding a Member of a
:
: TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
:
: NUMBER OF SEQUENCES: 14
:
: CORRESPONDENCE ADDRESSES:
:
: ADDRESSEE: Campbell and Flores
:
: STREET: 4370 La Jolla Village Drive, Suite 700
:
: CITY: San Diego
:
: STATE: California
:
: COUNTRY: USA
:
: ZIP: 92122
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/486,196
:
: FILING DATE:
:
: CLASSIFICATION: 424
:

```

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/118,906
 FILING DATE: 09-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LT 9526
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1807 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 255..1454
 US-08-486-196-13

Query Match 9.6%; Score 180.4; DB 1; Length 1807;
 Best Local Similarity 54.3%; Pred. No. 7.3e-49;

Matches 395; Conservative 0; Mismatches 321; Indels 12; Gaps 1;

388 tgtgagcaactcaaggctgaagaagttatcaatccactgagcaagaagtg 447
 468 tgcaggaataacttgaccagcagcactacatccaccccttattgaagaagagct 527
 448 gaattcccttgcatactatggtgatcagagaagattgaacattgaaggcta 507
 528 GACTTTCCTTGCGATATATATGTCATCATCATCTTGACCTTGACCTTGCAAGCTC 587
 508 ctgcagagctgtatagccctcagaacatatactgttcatgttgatgagtgagtcacca 567
 588 TTCAGGGCTTTTACATGCCCCCAATAATCTACTGTGTATGATGGAATGAAGACACA 647
 568 gaaacttcaagagcgcgcaagaacattatctgtctcccaatgtctcatagcc 627
 648 ACTGAATTTAAGATGCGGTAGAGCACTATTAAAGCTGCTCCCAAGCGTTTCTGCGCT 707
 628 agtaagctgtgtcggtgttatgtcctcctgtgcagagtgcaagctgcactcaatgc 687
 708 TCCAAAGATGGAACCCGCTTGTATGAGGGATCTCCAGGCTGCACTGAACCTGC 767
 688 atggaagactgtctccagagctcagtgccgttgaaatactccgaatcatatgtggagc 747
 768 ATCAGAGATCTTCTGCTTCGAGGTCTCATGGAAGTACGTATCAACACCTGTGGGCAA 827
 748 gacttccataaagcaatgcaagatgtgtcagagctcctcaagatgttgaatggaag 807
 828 GACTTCCCGCTGAACAACCAAGAAATAGTTACATATGAAAGATTTAAAGGTAA 887
 808 aatagcatgtgagcagagtgactcctcctacacaaagaacccgcgtggaatat----- 861
 888 AATATACCCCAAGGGGTGCTGCCCCAGCTCATGCAATTGAGAGACTAATATGTGCAC 947
 862 -----cactttagagtagagagacacattacactaacctaacagaagaagatcct 915
 948 CAAGAGCACTTGAGCAAGAGCTTCTTATGTGATTAAGAAACAAGCGGTGAACCGCT 1007
 916 cccctataatlaactatgattacaggaatgagctacattgtgctccagatctc 975
 1008 CCCCCCATATATCATCAATTTACTTTGCTGCTGCTCATGTGCTCATTAAGAGATTT 1067
 976 gtccaacatgttttgaagaaccttaatcccaacactgttgaatggtgaagacact 1035
 1068 GCCAACTTTTGTTCGTGCTGACCCAGCGGCTGTGATTTGCTCCAGTGGTCAAGGACCT 1127
 1036 tatagcccaatgaacaccccttgccaccccttcaggtgacagtgagctgagctc 1095

Db 1128 TTCAGTCCATGATGACATTTCTGGGTGACACTCAATAGATTCAGGTGTTCTGCTCT 1187
 QY 1096 gtcccaaa 1103
 Db 1188 ATGCCAAA 1195

RESULT 9

US-08-488-135-13
 Sequence 13, Application US/08488135
 Patent No. 5766910
 GENERAL INFORMATION:
 APPLICANT: Fukuda, Minoru
 APPLICANT: Bierhuizen, Marti F.A.
 TITLE OF INVENTION: Expression of the Developmental I
 TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell and Flores
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/488,135
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/118,906
 FILING DATE: 09-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LT 9526
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1807 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 255..1454
 US-08-488-135-13

Query Match 9.6%; Score 180.4; DB 1; Length 1807;
 Best Local Similarity 54.3%; Pred. No. 7.3e-49;

Matches 395; Conservative 0; Mismatches 321; Indels 12; Gaps 1;

388 tgtgagcaactcaaggctgaagaagttatcaatccactgagcaagaagtg 447
 468 tgcaggaataacttgaccagcagcactacatccaccccttattgaagaagagct 527
 448 gaattcccttgcatactatggtgatcagagaagattgaacattgaaggcta 507
 528 GACTTTCCTTGCGATATATATGTCATCATCATCTTGACCTTGACCTTGCAAGCTC 587
 508 ctgcagagctgtatagccctcagaacatatactgtgtccatgttgatgagaagtcacca 567
 588 TTCAGGGCTTTTACATGCCCCCAATAATATCTACTGTGTGATGATGGAAGAAAGACACA 647

QY	568	gaacattcaaaagagcggtgcacaaagaatattcttgcttccaaatgcttccaaagc	627
Db	648	ACTGAAATTTAAAGATGGGGAGAGCACTATTAAAGCTGCTTCCCAAAAGCTTTCTGGCT	707
QY	628	agtaagcttggtcggtggtgttaagcttcctcctgctccaggtgycgaagctgacctaaacgc	687
Db	708	TTCCAAAGTGGAAACCCGTTGTCATGGAAGGAGATCTCCAGGCTCCAGGCTGACCTGAACGC	767
QY	688	atggaagacttgctccagagctcagtgccgttggaataattcctcgtatatactgtggagcg	747
Db	768	ATCAGAAATTTCTCTGCTTCGAGGTCATGAGTACGTATCAACACCTGTGGGCAA	827
QY	748	gaacttctcataaagaagcaatgcaagatggtgcgcagcctctaagaatgttgaatggagcg	807
Db	828	GACTTCCCTCGAAACCAACGAAGAAATGTCTAGTATCTGAAGATTTAAAGGTAA	887
QY	808	aatagcattgagtcagagagtaacctccataagcacaagaacccgcttggaatat-----	861
Db	888	AATATCAACCCAGGGGGTGGTCCCCCAGCTCATGTGCAATTGGACGGACATTAATATGTCCAC	947
QY	862	-----cactltgaggtagtgagagacacattacctaactaaacaagaagaagatcct	915
Db	948	CAGAAGACCTGGGGCAAAAGAGCTTTCTATGTGATPAGAACAAACAGCTTGAACCCGCT	1007
QY	916	ccccctataatttaactatglttlaaaggaatgctacatgtgtcctcccgagatttc	975
Db	1008	CCCCCCCATATATCTCACAAAATTTACTTTGGCTGCTCGCTATGTGGCTCTATCAAGAGAGTTT	1067
QY	976	gtcccaacatgttlttgaagaacccttaactccaacaacatgatgaaatgggtlaaagaacct	1035
Db	1068	GCCAACTTTGTCTGCATGACCCACGGGCGTGTGATTTTGCTCCAATGGTCCAAAGCAACT	1127
QY	1036	tatagcacagatgaacaacctcttgggcacaccttcagcgtgcacggttggaatgcttgacct	1095
Db	1128	TTTCAGTCTCATGATGACATTTCTGGTGACACATCAATAGATTTCCAGGTGTTCTGGGCTCT	1187
QY	1096	gtttcccaa 1103	
Db	1188	ATGCGCAA 1195	

1 RESULT 10
 2 US-08-474-065-13
 3 Sequence 13 Application US/08474065
 4 Patent No. 5830465
 5 GENERAL INFORMATION:
 6 APPLICANT: Fukuda, Minoru
 7 APPLICANT: Bierhuizen, Marti F.A.
 8 TITLE OF INVENTION: Expression of the Developmental I
 9 TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
 10 NUMBER OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
 11 NUMBER OF SEQUENCES: 14
 12 CORRESPONDENCE ADDRESS:
 13 ADDRESSEE: Campbell and Flores
 14 STREET: 4370 La Jolla Village Drive, Suite 700
 15 CITY: San Diego
 16 STATE: California
 17 COUNTRY: USA
 18 ZIP: 92122
 19 COMPUTER READABLE FORM:
 20 MEDIUM TYPE: Floppy disk
 21 COMPUTER: IBM PC compatible
 22 OPERATING SYSTEM: PC-DOS/MS-DOS
 23 SOFTWARE: PatentIn Release #1.0, Version #1.25
 24 CURRENT APPLICATION DATA:
 25 APPLICATION NUMBER: US/08/474,065
 26 FILING DATE:
 27 CLASSIFICATION: 424
 28 PRIOR APPLICATION DATA:
 29 APPLICATION NUMBER: US 08/118,906
 30 FILING DATE: 09-SEP-1993
 31 ATTORNEY/AGENT INFORMATION:
 32 NAME: Campbell, Cathryn A.

```

;
; REGISTRATION NUMBER: 31,815
;
; REFERENCE/DOCKET NUMBER: P-LJ 9526
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (619) 535-9001
;
; TELEFAX: (619) 535-8949
;
; INFORMATION FOR SEQ ID NO: 13:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 1807 base pairs
;
; TYPE: nucleic acid
;
; STRANDEDNESS: double
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: cDNA
;
; FEATURE:
;
; NAME/KEY: CDS
;
; LOCATION: 255..1454
;
US-08-474-065-13

```

[illegible]

Db 133 CATGTGACACAAAATCCGAGATTCCTATTAGCTGCAGTGGCATCCCTTCCTGT 192
Qy 607 ttcccaaatgtctcatagaccagtaagctggtcggtgtttatagccctcgtgcccag 666
Db 193 TTTATTAATGCTTTGTGGCCAGCGATTGGAGAGTGTGTTATGCAATCGTGGACCGG 252
Qy 667 gtgcagctgaccctcaactgcatggaagactgtctccagagctcagtgccgtggaatac 726
Db 253 GTTCAGGCTGACCTCACTGATGAAAGATCTCTATGCAATGAGTGCAAACTGGAAGTAC 312
Qy 727 ttccgaatacatgtggtgagcgacttctctataaagagcaatgcagagatgtgccagct 786
Db 313 TTGATTAATCTTTGTGTGATGATTTTCCCATTAACCACTGAAATTTGTCAAGAG 372
Qy 787 ctcaag 792
Db 373 CTCAG 378

RESULT 13

US-08-488-135-3
Sequence 3, Application US/08488135
Patent No. 5766910

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru
APPLICANT: Bierhuizen, Marti F.A.
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488.135
FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378

US-08-488-135-3

Query Match 8.6%; Score 162.8; DB 1; Length 378;
Best Local Similarity 65.3%; Pred. No. 1.5e-43;
Matches 239; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 427 ccactgagaagaagaggtgagttccctatgcatactctatgtgtaatcatgagaag 486
Db 13 CCCCTTAGTAAGAAAGAGCGGAGTTTCCAAATAGCAATATTAAGTGGTTCATCAAG 72
Qy 487 attgaaacttggaaagctactcgagactgtgatgcccctcagaacatatactgtgc 546
Db 73 ATTGAATCTTGTGACAGGCTGCTGAGGGCCATCATATATCCCTGAGAAATTTATTCGGT 132
Qy 547 catgtgagatgaagttccccaagaacttccaagaaggcggtcaagaactatctctgc 606
Db 133 CATGTGACACAAAATCCGAGATTCCTATTAGCTGCAGTGGCATCCCTTCCTGT 192
Qy 607 ttcccaaatgtctcatagaccagtaagctggtcggtgtttatagccctcgtgcccag 666
Db 193 TTTATTAATGCTTTGTGGCCAGCGATTGGAGAGTGTGTTATGCAATCGTGGACCGG 252
Qy 667 gtgcagctgaccctcaactgcatggaagactgtctccagagctcagtgccgtggaatac 726
Db 253 GTTCAGGCTGACCTCACTGATGAAAGATCTCTATGCAATGAGTGCAAACTGGAAGTAC 312
Qy 727 ttccgaatacatgtggtgagcgacttctctataaagagcaatgcagagatgtgccagct 786
Db 313 TTGATTAATCTTTGTGTGATGATTTTCCCATTAACCACTGAAATTTGTCAAGAG 372
Qy 787 ctcaag 792
Db 373 CTCAG 378

RESULT 14

US-08-474-065-3
Sequence 3, Application US/08474065
Patent No. 5830465

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru
APPLICANT: Bierhuizen, Marti F.A.
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474.065
FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA

FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-474-065-3

Query Match 8.6%; Score 162.8; DB 2; Length 378;
Best Local Similarity 65.3%; Pred. No. 1.5e-43;
Matches 239; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 427 ccactgagcaagaaggggaggtccctatgtacatactctatggtatcatgagag 486
DB 13 CCCCTGTGTAAGAGAGGCGGAGTTTCCATAGCATATTTCTAATGCTTCAACAG 72
QY 487 attgaaccttgaaagagctactgagctgtgtatgccccctcagaacatactgtgc 546
DB 73 ATTGAATGCTTGACAGGCTCTGAGGCCATCTATATGCTTCAGAAATTTCTATGCGTT 132
QY 547 catgtgatgagaagtcgcccaaaacttcaagaaggcggtcaagaacataattcttgc 606
DB 133 CATGTGACACAAATCCGAGAGTTCTATTTAGCTGACAGATGGGATGCTTCTGT 192
QY 607 ttcccaatgcttcataagccagtaagctgtgtgtgtgtatgtcctcctgtccag 666
DB 193 TTAGTATGCTTTGTGGCCAGCCGATGGAAGTGTGTATGCTGTGAAGCCGG 252
QY 667 gtgcaagctgacccaactgcatgaaagactgtctccagaagctcgtcggaataac 726
DB 253 GTTACGCTGACCTCAACTGCATGAGGATCTATGCAATGAGTGCNAATGGAAGTAC 312
QY 727 ttcttgatacatgtggagagacttccctataaagaagcaatgcagagatgtccagct 786
DB 313 TTGTAATATCTTTGTGTATGATTTTCCATTAACCAACTGAATTTGTGAGAG 372
QY 787 ctcaag 792
DB 373 CTCAG 378

RESULT 15

US-08-118-906-1
Sequence 1, Application US/08118906
Patent No. 5484590
GENERAL INFORMATION:
APPLICANT: Fukuda, Minoru
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
TITLE OF INVENTION: Beta1,6-N-Acetylglucosaminyltransferase Gene Family
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,906
FILING DATE: 09-SEP-1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-118-906-1

Query Match 7.1%; Score 134; DB 1; Length 378;
Best Local Similarity 60.4%; Pred. No. 4.5e-34;
Matches 221; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 426 ccactgagcaagaaggggaggtccctatgtacatactctatggtatcatgagaa 485
DB 12 CCCCTTATCTTAAGAGAGAGCTGACTTCCCTTGCAATATTAATGCTTCATCATCA 71
QY 486 gattgaaccttgaagagctactgagctgtgtatgccccctcagaacatactgtgt 545
DB 72 CTTTGACACCTTTGCAAGGCTTTCAGGGCTATTTACATGCCCCAATATCTCTGT 131
QY 546 ccactgtgatgagaagtcgcccaaaacttcaagaaggcggtcaagaacataattcttgc 605
DB 132 TCATGTGATGTAAGAACCACTGAATTTAAGATGCGGTAGAGCAACTATTAAGCTG 191
QY 606 ctcccaaatgtcttcataagcaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 665
DB 192 CTTCCCAAGCCTTTTCTGTGCTCCAAAGTGAACCCGTGTCTATGAGGAGTCTCAG 251
QY 666 gtgcaagctgacccaactgcatgaaagactgtctccagaagctcgtcggtggaata 725
DB 252 GCTCAGGCTGACCTGAACTGCATGAGAGCTTTGCTTGCCTTGAGAGTCTCATGGAAGTA 311
QY 726 ctcttgatacatgtggagagacttccctataaagaagcaatgcagagatgtccagc 785
DB 312 CGTTATCAACACCTGTGGCAAGACTTCCCTGAAAAACCAAGAAATAGTTCACTA 371
QY 786 tctcaa 791
DB 372 TCTGAA 377

Search completed: September 26, 2002, 04:25:42
Job time: 9072 sec

THIS PAGE BLANK (U3PT0)

[illegible]

XX New nucleic acid molecules of core 2
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 PT compositions for treatment of disorders mediated by the enzyme
 PT including cancer, cardiovascular and inflammatory disorders.
 XX
 PS Claim 3; Page 50-51; 66pp; English.
 XX
 CC The present sequence represents a human core 2
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
 CC The polypeptide can be used to treat diseases and disorders, such as
 CC cancer, cardiovascular disorders and inflammatory disorders including
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as
 CC atherosclerosis and clotting can also be treated. The polypeptides of
 CC the invention are predominantly expressed in gastrointestinal tissue
 CC (stomach, colon, intestine, testis) and are elevated in cancer.
 CC gastrointestinal disorders that may be prevented or treated include
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 CC and ulcerative colitis. The antibodies may be used in
 CC immuno-histochemical analysis, to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular
 CC locations and to quantitate the level of expression.
 XX
 SQ Sequence 438 AA;

Query Match 100.0%; Score 2360; DB 21; Length 438;
 Best Local Similarity 100.0%; Pred. No. 2.3e-224;
 Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWQKRRLCOLHYLMALGCVMLATVALKLSFRKCDSDHILGLESRSOSQYCRNIIYNFL 60
 DB 1 mvgwkrllcqghylwalgymllatvalklsfrkcdshlglesresgqycnllynfl 60
 QY 61 KLPKRSINCSGVTRGDOEAVLOAILNNELEVKKKRPFDTHTYLSLRDCEHFKARKEFI 120
 DB 61 klprkrsinscgytrgdgeavllqailnnlevkkkrpfdthtylsltrdcehfkærkfi 120
 QY 121 QPFLSKEEVEFPFAYSMVHEKTEINFERLIRAYAPONTICVAVDEKSPTEFEAVKATII 180
 DB 121 qflskseevfplaysmvhiekiefenferlirayaponticvavdeksptefkæavkati 180
 QY 181 SCFPNFIASKLVRVYASWSRVQADLNCMEDLLQSSVPMKYPLNCGTDFPIKSAEMV 240
 DB 181 scfpnvfiasklvrvyaswsrvqadlnmedllqssvpmkyplntcgdfpiksaemv 240
 QY 241 QALKMLNGRMSMESEVPKPKETRMKTHFRVAVDTLHTRKKKDDPPYNTLTMTGNAYIV 300
 DB 241 qalkmlngrmsmesevpkketrmkthfrvavdtlhtltnkkddppynltmtgnayiv 300
 QY 301 ASRDFPQVHVLKNPKSOQLIEMVKTYSPPDEHLMATLQARAMPGSVNHFKYDISDWTSTI 360
 DB 301 asrdfpqnvlknpsqqliemvktysppdehlmatlqarampgsvnhpkysidwtsti 360
 QY 361 ARLVKKQGHGIDKGAAPACSGIHQRAICVYGAGDLNMMNLONHLLANKFDPKVDNA 420
 DB 361 arlvkkgghgdidkgaapacsgihqraicvygagdlnmmlonhllankfdpkvdna 420
 QY 421 LOCLEFLRYKATYGTGL 438
 DB 421 lqcleelyrkaitygtel 438

RESULT 2
 AA94492
 ID AA94492 standard; Protein; 438 AA.
 XX
 AC AA94492;
 XX
 DT 19-SEP-2000 (first entry)
 XX

DE Human C2/4Gnt protein.
 XX
 KW Human; C2/4Gnt; UDP-N-acetylglucosamine; O-glycan biosynthesis;
 KW O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 10..27
 XX /note="putative transmembrane domain"
 XX
 PN MO200034449-A2.
 XX
 XX 15-JUN-2000.
 XX
 PD 03-DEC-1999; 99WO-DK00677.
 XX
 PF 04-DEC-1998; 98DK-0001605.
 XX
 PR (CLAU/) CLAUSEN H.
 XX
 PA Clausen H, Schwientek T;
 PI WPI; 2000-423407/36.
 XX
 DR N-PSDB; AAA48623.
 XX
 PT New nucleic acid molecule encoding UDP-N-acetylglucosamine useful as
 PT probe for the detection of specified glucosaminyltransferase from
 PT other species and related organisms
 PS Claim 20; Fig 2; 47pp; English.
 XX

The present sequence is human UDP-N-acetylglucosamine:
 C2/4Gnt. C2/4Gnt is the third member of the family of O-glycan
 beta-1,6-N-acetylglucosaminyltransferases. It adds core 2 or core 4
 based O-glycans on to oligosaccharides, glycoproteins and
 glycosphingolipids. C2/4Gnt can therefore be used in the production of
 appropriately glycosylated glycoconjugates with particular enzymatic,
 immunogenic, or other biological or physical properties. The nucleotide
 sequence is useful as a probe for the detection of C2/4Gnt from other
 species and related organisms and for the recombinant production of
 C2/4Gnt polypeptide. The nucleotide sequence was identified by analysis
 of EST database sequence information. Oligonucleotides derived from EST
 clone 17865 of ATCC were used to isolate two full-length C2/4Gnt clones
 from a human foreskin genomic plasmid library by 5' RACE PCR. RT-PCR was
 performed using Coloz05 human cell line mRNA in order to produce cDNA
 for expression of C2/4Gnt in Sf9. The control of O-glycan core assembly
 has been implicated in tumour progression and metastasis.

Sequence 438 AA;

Query Match 100.0%; Score 2360; DB 21; Length 438;
 Best Local Similarity 100.0%; Pred. No. 2.3e-224;
 Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWQKRRLCOLHYLMALGCVMLATVALKLSFRKCDSDHILGLESRSOSQYCRNIIYNFL 60
 DB 1 mvgwkrllcqghylwalgymllatvalklsfrkcdshlglesresgqycnllynfl 60
 QY 61 KLPKRSINCSGVTRGDOEAVLOAILNNELEVKKKRPFDTHTYLSLRDCEHFKARKEFI 120
 DB 61 klprkrsinscgytrgdgeavllqailnnlevkkkrpfdthtylsltrdcehfkærkfi 120
 QY 121 QPFLSKEEVEFPFAYSMVHEKTEINFERLIRAYAPONTICVAVDEKSPTEFEAVKATII 180
 DB 121 qflskseevfplaysmvhiekiefenferlirayaponticvavdeksptefkæavkati 180
 QY 181 SCFPNFIASKLVRVYASWSRVQADLNCMEDLLQSSVPMKYPLNCGTDFPIKSAEMV 240
 DB 181 scfpnvfiasklvrvyaswsrvqadlnmedllqssvpmkyplntcgdfpiksaemv 240

OY 241 QALKMLNGRNSMSEVPPKHKETRMKYHFEVVRDTLHLTNKKKDDPPYNLTFMTGNAYIV 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 241 galikmlngnrsmesevppkhetrkwkyhfevvrtdlhltnkkkddppynltmfignayiv 300
 OY 301 ASRDFVQVHLKMPKSOQLIEWKDYTSPEDEHLMATLQARMMGSGVNPHPKDYDISDMSI 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 301 asrdfvqvhvlnkpksgqliewkdytspeheltwatlqarwmpgsvnpnphkydisdmsi 360
 OY 361 ARLVWMOGHEGIDKGPAPCSGIHQARICVYGAGDLNMMONHLLANKRDPKYDDNA 420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 361 arlvwmghegidkgapapcsghlqaricvysagdlmmiqnhllankrtdpkvddna 420
 OY 421 LQCLEEYLRKAIYGTETL 438
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 421 lqcleeylrkaiygtetl 438
 RESULT 3
 AAB30518 standard; Protein: 438 AA.
 XX AAB30518;
 AC AAB30518;
 DT 06-MAR-2001 (first entry)
 DE Amino acid sequence of beta-1-6-N-acetylglucosaminyltransferase.
 XX
 KW Human: beta-1-6-N-acetylglucosaminyltransferase; C2GNT-M; Inflammation;
 KM membrane protein; branched sialyl Lex; L-selectin; immune reaction;
 XX Inflammation; tissue rejection; tumour metastasis.
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT Modified-site 289
 FT /note- "potential N-glycosylation site"
 XX
 PN US6136580-A.
 XX
 PD 24-OCT-2000.
 XX
 PE 19-JAN-1999; 99US-0233506.
 XX
 PR 19-JAN-1999; 99US-0233506.
 XX
 PA (BURN-) BURNHAM INST.
 XX
 PI Fukuda M, Yeh J;
 XX
 DR WPI: 2001-040238/05.
 DR N-PSDB: AAC62134.
 XX
 PT New C2GNT-M polypeptides having core 2, core 4 and I branching
 PT beta-1-6-N-acetylglucosaminyltransferase activities for preparing
 PT reagents useful for diagnosing, preventing or treating inflammation or
 PT tumour metastasis -
 XX
 PS Claim 1; Fig 4; 25pp; English.
 XX
 CC The present sequence represents a human
 CC beta-1-6-N-acetylglucosaminyltransferase polypeptide, with core2, core4
 CC and I branching activities. It is designated C2GNT-M. C2GNT-M is a
 CC membrane protein that is predominantly expressed in colon, small
 CC intestine, trachea, stomach and thyroid, as well as in certain cancer
 CC cell lines. C2GNT-M polypeptides may be used to prepare molecules having
 CC highly branched sialyl Lex and L-selectins, which may be subsequently
 CC used to modulate immune reactions, e.g. inflammation and tissue
 CC rejection, and to prevent or inhibit tumour metastasis.
 XX
 SQ Sequence 438 AA;

Best Local Similarity 100.0%; Pred. No. 2,36-224;
 Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 MYQWRKLCQLHLNMLAGCMMLATYALKSFPLKCDSDHLLGLESRSSQCYRNLINYL 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 mwqwrklcqlhlmlagcmmlatyalksfplkcdsdhllglessrsgqcrnllnlyl 60
 OY 61 KLPKRSINCSGVTGDDQAVLQALINLNLEVKRREPFDTHYLSLRDCEHFKAERKFI 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 klprksinscsvtrgddqavvlqalnlleevkrrepfdthylslrdcehfkaerf 120
 OY 121 QPPLSKEVEYFPDIAYSMVTHEKTIENFERLLRAVYAPQNIYCVHDEKSETEKAVKALI 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 121 qpllskeveyfpdiaysmvthehtienferllravypqnlycvhdeksefekaavk 180
 OY 181 SCFPNWFIAASKLVRYVYASMSRYQADLNCMEDLQSSVWPKFPLNTCGDFFIKSAENV 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 181 scfpnwfiasklvrvyvasmsryqadlncmmdlqssvwpkyflntcgdfpiksnaenv 240
 OY 241 QALKMLNGRNSMSEVPPKHKETRMKYHFEVVRDTLHLTNKKKDDPPYNLTFMTGNAYIV 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 241 galikmlngnrsmesevppkhetrkwkyhfevvrtdlhltnkkkddppynltmfignayiv 300
 OY 301 ASRDFVQVHLKMPKSOQLIEWKDYTSPEDEHLMATLQARMMGSGVNPHPKDYDISDMSI 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 301 asrdfvqvhvlnkpksgqliewkdytspeheltwatlqarwmpgsvnpnphkydisdmsi 360
 OY 361 ARLVWMOGHEGIDKGPAPCSGIHQARICVYGAGDLNMMONHLLANKRDPKYDDNA 420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 361 arlvwmghegidkgapapcsghlqaricvysagdlmmiqnhllankrtdpkvddna 420
 OY 421 LQCLEEYLRKAIYGTETL 438
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 421 lqcleeylrkaiygtetl 438
 RESULT 4
 AAG75058 standard; Protein: 465 AA.
 XX AAG75058;
 AC AAG75058;
 DT 03-SEP-2001 (first entry)
 DE Human colon cancer antigen protein SPQ ID NO:5822.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KM colorectal carcinoma; chromosome 15.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PE 28-SEP-2000; 2000WO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX
 DR WPI: 2001-235357/24.
 DR N-PSDB: AAH34463.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 PS Claim 11; Page 7338-7340; 9803pp; English.
 XX
 AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

Query Match

100.0%; Score 2360; DB 22; Length 438;

CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 CC
 XX Sequence 465 AA:

Query Match 100.0%; Score 2360; DB 22; Length 465;
 Best Local Similarity 100.0%; Pred. No. 2,6e-224;
 Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWQMKRLCOLHYLMALGCMYLATVALKLSFRKCDSDHGLGESRESOSQYCRNLTNPL 60
 DB 28 mwqkrlcqlhylwalgcmlylatvalklksfrlkcdsdhglgesresqycrnltynfl 87
 QY 61 KLPKRSINSGVTRGQOEAVLQAIINNLEKKKREPTDTHYSLTRDCEHFAERKFI 120
 DB 88 klprkrsinsgytrgqeaavlqalinnlekkkrepftdthysltrdcehfaerkfii 147
 QY 121 QPFLSKVEVEPIAYSWIHEKIEFRLRAYAPONICVHDEKSPETFKFAVAII 180
 DB 148 qflpskvevefpiayswihelkiefrrlayapqnlycvhdekspefkfaavaii 207
 QY 181 SCFPNVFIASKLVNVYASRSVQADNCMEDLLQSSVPKRYFLNTGTFPIKSNAEMV 240
 DB 208 scfpnvfiasklvrvyvasrsvqadnccmedllqssvprkyflntgctfpiksnaemv 267
 QY 241 QALKMLNGRSMSEVPKPKETRMKYHFEVVRDTLTLTKKKRPPRYNLTMFGNAIY 300
 DB 268 qalkmlngrsmesevprpketrmkyhfevvrdtlhlnkkkrrpprynlmtfgnaiyiv 327
 QY 301 ASRDFVOHVLKNRKSQOLIMVVDIYSPDEHLMATLQARWMPGSPVNHPRKYDISDMTSI 360
 DB 328 asrdfvohvlnkrksqolimvvdityspdehlmatlqarwmpgspvnhprkydisdmtsi 387
 QY 361 ARLYKMGEGDIDKAPYAPCSIGHORATCVGAGDLNMMLOHNLANKPDKVDNA 420
 DB 388 arlykmgegdldkapyapcsgihoratcvgagdlnmmlonhnlankpdkvddna 447
 QY 421 LQCLEEYLRYKATYGTGL 438
 DB 448 lqcleeylrykaygtel 465

RESULT 5
 AAB18996
 ID AAB18996 standard; Protein; 663 AA.

XX AAB18996:
 XX 08-FEB-2001 (first entry)
 XX
 DE A partial core 2 beta-1,6-N-acetylglucosaminyltransferase.
 XX
 XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
 KW cancer; cardiovascular disorder; inflammatory bowel disease; arteriosclerosis;
 KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 KW septic shock; adult respiratory distress syndrome; ARDS; cancer;

KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KW diverticulitis; ulcerative colitis.

OS Homo sapiens.

PN CA2296936-A1.

PD 03-AUG-2000.

PF 03-FEB-2000; 2000CA-2296936.

PR 03-FEB-1999; 99US-0118674.

XX (GLYC-) GLYCODESIGN INC.

PI Koczak B, Lew A;

DR WPI; 2000-594746/57.

PT New nucleic acid molecules of core 2

PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new

PT compositions for treatment of disorders mediated by the enzyme

PT including cancer, cardiovascular and inflammatory disorders.

PS Claim 3; Page 52; 66pp; English.

CC The present sequence represents a partial human core 2
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
 CC The polypeptide can be used to treat diseases and disorders, such as
 CC cancer, cardiovascular disorders and inflammatory disorders including
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as
 CC atherosclerosis and clotting can also be treated. The polypeptides of
 CC the invention are predominantly expressed in gastrointestinal tissue
 CC (stomach, colon, intestine, testis) and are elevated in cancer.
 CC Gastrointestinal disorders that may be prevented or treated include
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 CC and ulcerative colitis. The antibodies may be used in
 CC immuno-histochemical analysis, to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular
 CC locations and to quantitate the level of expression.

XX Sequence 663 AA:

Query Match 100.0%; Score 2360; DB 21; Length 663;
 Best Local Similarity 100.0%; Pred. No. 4,5e-224;
 Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWQMKRLCOLHYLMALGCMYLATVALKLSFRKCDSDHGLGESRESOSQYCRNLTNPL 60
 DB 85 mwqkrlcqlhylwalgcmlylatvalklksfrlkcdsdhglgesresqycrnltynfl 144
 QY 61 KLPKRSINSGVTRGQOEAVLQAIINNLEKKKREPTDTHYSLTRDCEHFAERKFI 120
 DB 145 klprkrsinsgytrgqeaavlqalinnlekkkrepftdthysltrdcehfaerkfii 204
 QY 121 QPFLSKVEVEPIAYSWIHEKIEFRLRAYAPONICVHDEKSPETFKFAVAII 180
 DB 205 qflpskvevefpiayswihelkiefrrlayapqnlycvhdekspefkfaavaii 264
 QY 181 SCFPNVFIASKLVNVYASRSVQADNCMEDLLQSSVPKRYFLNTGTFPIKSNAEMV 240
 DB 265 scfpnvfiasklvrvyvasrsvqadnccmedllqssvprkyflntgctfpiksnaemv 324
 QY 241 QALKMLNGRSMSEVPKPKETRMKYHFEVVRDTLTLTKKKRPPRYNLTMFGNAIY 300
 DB 325 qalkmlngrsmesevprpketrmkyhfevvrdtlhlnkkkrrpprynlmtfgnaiyiv 384
 QY 301 ASRDFVOHVLKNRKSQOLIMVVDIYSPDEHLMATLQARWMPGSPVNHPRKYDISDMTSI 360
 |||||||

Db 365 asdfvghvlnkpsqqllewktdyspdehlwatlqrarwmpgsvphkpydtsdmtsi 444
 Qy 361 ARLVKMOGHEDIDKGAAPYAPCSGIHORAICVYGAGDINMMLONHHLANKRDPKVDNA 420
 Db 445 arlvkmghegdidkgayapcsghraicvysgadinmnlqnhllankfdpkvdna 504
 Qy 421 LQCEEYLRKAIYGTTEL 438
 Db 505 lqcleeylrkaiygtel 522

RESULT 6
 AAB54344
 ID AAB54344 standard; Protein: 465 AA.
 AC AAB54344;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human pancreatic cancer antigen protein sequence SEQ ID NO:796.
 XX
 KW Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW nocotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiac; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative.

XX Homo sapiens.
 OS
 PN MO200055320-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000MO-US05989.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-579444/54.
 DR N-PSDB; AAC99109.
 XX
 PF New nucleic acid that is a pancreatic cancer antigen for preventing,
 PF treating, or ameliorating a medical condition, particular pancreatic
 PI cancer, or for use in assays for diagnosing a pathological condition -
 XX
 XX Claim 11; Page 1237-1239; 1379pp; English.

CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, nocotropic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiac and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent

CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 465 AA;
 Query Match 99.7%; Score 2354; DB 21; Length 465;
 Best Local Similarity 99.8%; Pred. No. 1e-223;
 Matches 437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MVQWKRICOLHYIMALGCMYMLATVALKLSPLKCDSDHGLGESRESOSOCRNILYPL 60
 Db 28 mvqwkrlcqlhyimalgcmmlatvalklsfklcdshlglstresgqcrnllynfl 87
 Qy 61 KLPAKRSINSGVTRGDPAVLQALINLLEVKKRREPFDTHYLSLTRDCEHFAERKEFI 120
 Db 88 klpakrsincsgvttrgdgavlgallmlevkkrreptfdthylsttrdcchfkerfki 147
 Qy 121 QPPLSKEEVERPIAYSMWYIHEKIEHFERLLRAVAPONIVCYHVDKSPETFEKAVKAI 180
 Db 148 qpllskeeverpiaysmwvihelkieferrllravapgnlycvhvdkspetfkeavkai 207
 Qy 181 SCFPPVFIASKLVRYVYASMSRVQADLNCMEDLQSSVPWKYELNWCGRDPPIKSAEMV 240
 Db 208 scfpnvfiasklvrvyaswsvrgadlnomedllqssvpwkyflntcgldfpiksaemv 267
 Qy 241 QALKMLNGRNSMESEVPKHKETRMKYHFEVVRDTLHLTNKKDDPPYNLTFGTGNAYIV 300
 Db 268 qalkmlngnrnsmevppkhketrkynhfevvrtdlhltnkkdppynlntfgtnayiv 327
 Qy 301 ASRDFVQHLVKNPKSQQLIEVWKDTYSPDEHLWATLQRRMMPGSPVPHPKYDISDMTSI 360
 Db 328 asdfvghvlnkpsqqllewktdyspdehlwatlqrarwmpgsvphkpydtsdmtsi 387
 Qy 361 ARLVKMOGHEDIDKGAAPYAPCSGIHORAICVYGAGDINMMLONHHLANKRDPKVDNA 420
 Db 388 arlvkmghegdidkgayapcsghraicvysgadinmnlqnhllankfdpkvdna 447
 Qy 421 LQCEEYLRKAIYGTTEL 438
 Db 448 lqcleeylrkaiygtel 465

RESULT 7
 AAB18999
 ID AAB18999 standard; Peptide: 406 AA.
 AC AAB18999;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE A core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region.
 XX
 KW Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
 KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
 KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 KW septic shock; adult respiratory distress syndrome; AIDS; cancer;
 KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KW diverticulitis; ulcerative colitis.
 KW
 XX
 XX Homo sapiens.
 OS
 PN CA2296936-A1.
 XX
 PD 03-AUG-2000.
 XX
 PF 03-FEB-2000; 2000CA-2296936.
 XX
 PR 03-FEB-1999; 99US-0118674.
 XX
 PA (GLYC-) GLYCDESIGN INC.
 XX
 PI Korczak B, Lew A;

XX WPI: 2000-594746/57.
 DR New nucleic acid molecules of core 2
 XX beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 PT compositions for treatment of disorders mediated by the enzyme
 PT including cancer, cardiovascular and inflammatory disorders.
 XX
 XX Claim 3; Page 53; 66pp; English.
 XX
 CC The present sequence is derived from a human core 2
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
 CC The polypeptide can be used to treat diseases and disorders, such as
 CC cancer, cardiovascular disorders and inflammatory disorders including
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as
 CC atherosclerosis and clotting can also be treated. The polypeptides of
 CC the invention are predominantly expressed in gastrointestinal tissue
 CC (stomach, colon, intestine, testis) and are elevated in cancer.
 CC Gastrointestinal disorders that may be prevented or treated include
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 CC and ulcerative colitis. The antibodies may be used in
 CC immuno-histochemical analysis, to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular
 CC locations and to quantitate the level of expression.
 XX
 SQ Sequence 406 AA:
 Query Match 92.6%; Score 2185; DB 21; Length 406;
 Best Local Similarity 100.0%; Pred. No. 4, 2e-207;
 Matches 406; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 33 LKCDSDHLGESHESQSOICRNILYNFLKPAKRSINCSGVTGRDQDAVQATINNEVK 92
 DB 1 lkcdshlgiesresgsgqgcrnllnflkpkarrsincsgvtgrdgaavqallnnevk 60
 QY 93 KRREPFDTHYLSLRDCEHFKAERKFIOPLSKEVEFPDIASMTHEKIENTERLLRA 152
 DB 61 kkrpfdthylsltrdcehfkaerfkifqlskevefpdiawvshkienterllra 120
 QY 153 VYAPQNTYCHVDEKSEPTKEAVKATISCFPRVFTASKIVRYVYASMSVQADLNKMD 212
 DB 121 vyapqnlcyhvdekspepkavkalisfprvftasklvyvvaswsvqadlnkmed 180
 QY 213 LLOSSVPWKYFLNTGDFPIKSNAEVQALKKMLNGRNSMESEVPKHKETRMKYHPEVY 272
 DB 181 llqssvpwkyflntgdfpiksnaevqalkkmlngrnsesevpkhketrmkyhpevy 240
 QY 273 RDTLHLTNKKKDDPPYNTLWFTGNAYIVASRDFVQHVHLNPKSQQLIEWVKDTYSPDEHL 332
 DB 241 rdtlhltnkkkddppynltwftgnayivaasrdfvqhvhlknpksgqliewvkdtyspdehl 300
 QY 333 WATLQARARWPGSVNHPKRYDIDMSIARLVWQGHGIDKGAIPAPSGIHQARICV 392
 DB 301 watlqararwpgsvnphkrydisdmsiarlvwqghgidkgaipapsgihqaricv 360
 QY 393 YGAGDINMMLONHHLANKRPDVNDNALOCLEELYRYKAIYGTML 438
 DB 361 ygagdlnmmlonhhlankrpdvndnalocleelyrykaiygtel 406
 RESULT 8
 AAM93943
 ID AAM93943 standard; Protein; 428 AA.
 XX
 AC AAM93943;
 XX
 DT 28-JUN-1999 (first entry)
 XX
 DE Human core 2 Gnt protein.
 XX

KW Screening; treatment; prevention; cardiomyopathy; inhibitor;
 KW diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;
 KW UDP-GlcNAc:Galbeta1-3GalNAc-alpha1; transgenic animal; germ line;
 KW beta-1,6-N-acetylglucosaminyl-1-transferase; human; core 2 Gnt.
 XX
 OS Homo sapiens.
 XX
 PN CA2186987-A.
 XX
 PD 02-APR-1998.
 XX
 PF 02-OCT-1996; 96CA-2186987.
 XX
 PR 02-OCT-1996; 96CA-2186987.
 XX
 PA (MOUN) MOUNT SINAI HOSPITAL CORP.
 XX
 PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE;
 XX
 DR WPI: 1998-399608/35.
 DR N-PSDB; AAX24043.
 XX
 PT Screening for substances that prevent or treat cardiomyopathy
 PT associated with diabetes and hyperglycaemia - comprises reacting
 PT core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor
 PT in presence of test substance
 XX
 PS Disclosure; Fig 9; 35pp; English.
 XX
 CC This invention describes a method for screening for a substance that
 CC may be used to prevent or treat cardiomyopathy associated with diabetes
 CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
 CC (UDP-GlcNAc:Galbeta1-3GalNAc-alpha1; beta-1,6-N-acetylglucosaminyl-
 CC transferase) with an acceptor substrate and a sugar nucleotide donor in
 CC the presence of a test substance under conditions whereby the core 2
 CC GlcNAc-T produces a reaction product, determining the amount of reaction
 CC product, and comparing the amount of reaction product with the amount
 CC obtained in the absence of the test substance, where lower amounts of
 CC reaction product in the presence of the test substance indicate that the
 CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)
 CC methods for preventing or treating cardiomyopathy associated with
 CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
 CC germ cells and somatic cells all contain a DNA construct introduced into
 CC the animal or an ancestor of the animal at an embryonic stage, where
 CC incorporation of the DNA construct into the germ line of the animal
 CC causes the animal to develop cardiomyopathy similar to that associated
 CC with diabetes mellitus and hyperglycaemia. This sequence represents the
 CC human core 2 Gnt protein (beta-1,6-N-acetylglucosaminyltransferase)
 CC which is used to describe the method of the invention.
 XX
 SQ Sequence 428 AA:
 Query Match 48.4%; Score 1142.5; DB 19; Length 428;
 Best Local Similarity 57.5%; Pred. No. 5, 4e-104;
 Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;
 QY 68 INCGVTRGDOEAVLOALINNEVK-KRREPFDTHYLSLRDCEHFKAERKFIOPLSK 126
 DB 57 inctkvlgdvneigkvkllvtckfkprwvppddyimtsccsfkikryiveplsk 116
 QY 127 EEVEFPDIASMTHEKIENTERLLRAVYAPQNTYCHVDEKSEPTKEAVKATISCFPRV 186
 DB 117 eeaeffdiaysivvhhkilemldrllraylmpqntycvhdtkesdsyaaavmgascsfnv 176
 QY 187 FIASKLVRYVYASMSRVQADLNKMDLQSSVPWKYFLNTGDFPIKSNAEVQALKKML 246
 DB 177 fvasrlsvyvaswsvqadlnmkdllyamsanwkylinlqgmfdpiktleivrkil 236
 QY 247 NGRNSMESEVPKHKETRMKYHPEVYRDTLHLTNKKKDDPPYNTLWFTGNAYIVASRDFV 306
 DB 237 mgenmletermpshkeerwkyevvngkltntgvlkmlpbletpflfsagayivsvreyv 296

QY 307 QHVLKPKSQQLIEWVKDTYSPDEHLWATLQARMMPGSVNHPKXIDISMTSIARLVKW 366
 Db 297 gylvlnqekigklmewagdqlyspdeylwatlgripevpslpashkydlsdmqavarfvkw 356
 QY 367 QGHGDDIDKGAAPYAPCSITHORATCYVGAGDLNMLQNHLLANKFDPKVDNALQCLIEE 426
 Db 357 qyfeqdvsksgapypcdqgvhvrsvclfgagdlmwllrkhlhlfankfdvdvdlfaigclde 416
 QY 427 YLRKAI 433
 Db 417 hlrhkal 423

RESULT 9
 AAB30298
 ID AAB30298 standard; Protein: 428 AA.
 AC AAB30298;
 DT 12-FEB-2001 (first entry)
 DE Human heart core 2 GlcNAc-T.
 KW Human: diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia;
 KW UDP-GlcNAc:Galbeta1-3GalNAcalpha1phar beta1-6
 KW N-acetylglucosaminyltransferase.
 OS Homo sapiens.
 PN US6131578-A.
 PD 17-OCT-2000.
 PE 02-OCT-1997; 97US-0943058.
 PR 02-OCT-1996; 96US-0046876.
 PA (KING/) KING G L.
 PA (NISH/) NISHIO Y.
 PA (KOYA/) KOYA D.
 PA (DENN/) DENNIS J W.
 PA (WARR/) WARREN C E.
 PI Nishio Y, Koya D, King GL, Warren CE, Dennis JW;
 DR WPI: 2000-678642/66.
 DR N-PSDB: AAC65469.
 XX Preventing or treating cardiomyopathy associated with diabetes mellitus
 PT and hyperglycemia, comprises administering a substance that inhibits
 PT core 2 N-acetylglucosamine-T activity -
 XX Claim 3; Fig 9; 21pp; English.
 CC The present invention describes the human UDP-GlcNAc:Galbeta1-3GalNAc
 CC alpha1phar beta1-6 N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein
 CC and coding sequences. The enzyme is associated with cardiomyopathy in
 CC diabetes and hyperglycemia sufferers. The invention also provides
 CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used
 CC in treatment.
 XX Sequence 428 AA;
 SQ

Query Match 48.4%; Score 1142.5; DB 21; Length 428;
 Best Local Similarity 57.5%; Pred. No. 5.4e-104;
 Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;

QY 68 INCSGVTGDDQEAVALQATLNNLEVK-KKREPTDTHTSLTRDCENHFAERKFTQFPLSK 126
 Db 57 lncktkvlgdvneiqkvlellvtkrkrptwtpddqymnmsdesfkkrrkyivspk 116
 QY 127 EEVEPPIAYSVIHEKIEFRLRLRAYAPONTICVHDEKSPETFEKAVAKIISCFPNV 186

Db 117 eeaeefpaysivvhhkilemldrlllraiympqafycvhdtskedsyaaavmgiascsfnv 176
 QY 187 FLASKLAVRVVYASWSRVQADLNCMEDLLQSSVPWKYFLNTGTFEPKISNEMVQALKML 246
 Db 177 fvasrlsvvyasvsvyqadlncmkdlyamsanwkyllnlcgmdprktnleivrkkl 236
 QY 247 NGRNSMESEVPKHKETFMWKHFEVVRDTLHLTNKKKDPPEYNLMFGNAYIVASRDFV 306
 Db 237 mgenmletermpshkeerwkkryevvngklntgtvkmplpplerplfsasayfvsvrey 296
 QY 307 QHVLKPKSQQLIEWVKDTYSPDEHLWATLQARMMPGSVNHPKXIDISMTSIARLVKW 366
 Db 297 gylvlnqekigklmewagdqlyspdeylwatlgripevpslpashkydlsdmqavarfvkw 356
 QY 367 QGHGDDIDKGAAPYAPCSITHORATCYVGAGDLNMLQNHLLANKFDPKVDNALQCLIEE 426
 Db 357 qyfeqdvsksgapypcdqgvhvrsvclfgagdlmwllrkhlhlfankfdvdvdlfaigclde 416
 QY 427 YLRKAI 433
 Db 417 hlrhkal 423

RESULT 10
 AAR51386
 ID AAR51386 standard; Protein: 428 AA.
 AC AAR51386;
 DT 07-OCT-1994 (first entry)
 DE Sequence of human core 2 beta 1-6
 DE N-acetylglucosaminyltransferase (C2GNT or core 1-6 AGT).
 DE C2GNT; 1-6 AGT; core 2 beta 1-6 n-acetylglucosaminyltransferase;
 KW O-glycan.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Domain 10..32
 FT Modified-site 52
 FT /label= signal/membrane anchoring domain
 FT /label= potential N-glycosylation site
 FT /note= "see also Aas 58 and 95"
 XX EP590747-A.
 XX 06-APR-1994.
 XX 29-SEP-1993; 93EP-0250268.
 XX 01-OCT-1992; 92US-0955041.
 XX (LJOL-) LA JOLLA CANCER RES FOUND.
 XX Blerhulzen MFA, Fukuda M;
 DR WPI: 1994-111195/14.
 DR N-PSDB: AAQ61539.
 PT New beta 1-6 N-acetylglucosaminyl transferase and acceptor - are
 PT used for the study of the effect of variant O-glycan(s) on
 PT cell-cell interactions, partic. in cancers
 XX Disclosure; Page 20-22; 34pp; English.
 CC C2GNT, or an active fragment thereof, catalyses the formation of
 CC critical branches in O-glycans. cDNAs encoding various
 CC glycosyltransferases can be isolated by transient expression of cDNA
 CC in recipient cells, e.g. COS-1. COS-1 cells were transfected
 CC with a cDNA library, pCDR alpha-2Fl, constructed from poly(A)+ RNA

CC of activated T lymphocytes which express the C2Gnt. Transfected
CC cells were selected using MAb T305, which identifies a
CC hexasaccharide on leukostallin. Leukostallin CD43 is an acceptor
CC molecule for C2Gnt activity. A plasmid, pCDR alpha-Leu,
CC which directed expression of the T305 antigen was identified. The
CC cDNA insert was isolated and sequenced. The cDNA encoded the
CC acceptor molecule leukostallin CD43.

XX
SQ Sequence 428 AA:

Query Match 47.9%; Score 1129.5; DB 15; Length 428;
Best Local Similarity 56.9%; Pred. No. 1.1e-102;
Matches 209; Conservative 57; Mismatches 100; Indels 1; Gaps 1;

QY 68 INCSGVTGDOEAVYQALINLNLEVK-KKRPFTDTHTLSLRDCEHFAERKFIQFPLSK 126
DB 57 INCTKVLGQDVNEIGKVKLEILTVKFKKRPWPDYIMTSCDSSFIKRPKYIVELPSK 116
QY 127 EEVEFFPIATSMVTHEKIEENFERLLRAVAPQNTYCVHDEKSPETREAVKAIISCFPNV 186
DB 117 EEEEFPIATSMVTHEKIEENFERLLRAVAPQNTYCVHDEKSPETREAVKAIISCFPNV 176
QY 187 FIASKLVRYVYASMSRVQADLNCMEDLQSSVPKRYFLNCGTDPPIKSNAEVQALKML 246
DB 177 FVASLIESVYASMSRVQADLNCMEDLQSSVPKRYFLNCGTDPPIKSNAEVQALKML 236
QY 247 NGRNSMESEVPRPKKETRMKRYHEVVRDTLHLTKRRKDPPEYMLTFPGNAVIVASRDFV 306
DB 237 MGSNLETETMPHKEERWKRYEYVNGKLTGTCTKMLPRLTCLPFIQSAGYFVTSREY 296
QY 307 QHVLKPKSQOQLIEWYKDTYSPDEHLMATLQARAKMPSVPRNHPKDISMTSIRLVK 366
DB 297 GYVLGQEKIKQKIMEWAGQDYSPEYIWTATQIPREVPSIPASHKYDLSDMGAVARFVK 356
QY 367 OGHESGIDIGAPYAPCSGIHQRAICVYGAGDLNMMLONHHLANKRPKVDNALQCLE 426
DB 357 GYFEGDVSKGAPYPPCSGVHVRSVCLFGAGDLNMMLRKHHLFANKTIDVDVLAIGQCLDE 416
QY 427 YLRKYAI 433
DB 417 HLRKAI 423

RESULT 11

AAM93942
ID AAM93942 standard; Protein; 428 AA.

AC AAM93942;

XX 28-JUN-1999 (first entry)

DE Rat DH1 protein.

KW DH1: rat; screening; treatment; prevention; cardiomyopathy; inhibitor;
KW diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;
KW UDP-GlcNAc:Galbeta1-3GalNAc-alpha; transgenic animal; germ line;
KW beta-1,6-N-acetylglucosaminyl-transferase.

OS Rattus sp.

PN CA2186987-A.

XX 02-APR-1998.

PF 02-OCT-1996; 96CA-2186987.

PR 02-OCT-1996; 96CA-2186987.

XX (MOON) MOUNT SINAI HOSPITAL CORP.

PI Dennis JW, King GL, Koya D, Nishio Y, Warren CE;

DR MPI: 1998-399608/35.
DR N-PSDB: AAX24042.

XX Screening for substances that prevent or treat cardiomyopathy
XX associated with diabetes and hyperglycaemia - comprises reacting
XX core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor
XX in presence of test substance

XX Example 1; Fig 3A; 35pp; English.

XX This invention describes a method for screening for a substance that
XX may be used to prevent or treat cardiomyopathy associated with diabetes
XX and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
XX (UDP-GlcNAc:Galbeta1-3GalNAc-alpha-beta-1,6-N-acetylglucosaminyl-
XX transferase) with an acceptor substrate and a sugar nucleotide donor in
XX the presence of a test substance under conditions whereby the core 2
XX GlcNAc-T produces a reaction product, determining the amount of reaction
XX product, and comparing the amount of reaction product with the amount
XX obtained in the absence of the test substance, where lower amounts of
XX reaction product in the presence of the test substance indicate that the
XX substance inhibits core 2 GlcNAc-T. The invention also describes (1)
XX methods for preventing or treating cardiomyopathy associated with
XX diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
XX germ cells and somatic cells all contain a DNA construct introduced into
XX the animal or an ancestor of the animal at an embryonic stage, where
XX incorporation of the DNA construct into the germ line of the animal
XX causes the animal to develop cardiomyopathy similar to that associated
XX with diabetes mellitus and hyperglycaemia. This sequence represents the
XX rat DH1 protein which is used to describe the method of the invention.

SQ Sequence 428 AA:

Query Match 46.9%; Score 1107.5; DB 19; Length 428;
Best Local Similarity 54.6%; Pred. No. 1.6e-100;
Matches 201; Conservative 64; Mismatches 102; Indels 1; Gaps 1;

QY 67 SINCSGVTGDOEAVYQALINLNLEVKKKRPFDTHTLSLRDCEHFAERKFIQFPLSK 125
DB 56 NVNCTKVLGQDVNEIGKVKLEILTVGFKKRPRTTPDHYIMTSCDSSFIRTKYIMPELT 115
QY 126 KEEVEFFPIATSMVTHEKIEENFERLLRAVAPQNTYCVHDEKSPETREAVKAIISCFPN 185
DB 116 KEEVGFPIATSMVTHEKIEENFERLLRAVAPQNTYCVHDEKSPETREAVKAIISCFPN 175
QY 186 VFIASKLVRYVYASMSRVQADLNCMEDLQSSVPKRYFLNCGTDPPIKSNAEVQALKML 245
DB 176 VFIASGLIESVYASMSRVQADLNCMEDLQSSVPKRYFLNCGTDPPIKSNAEVQALKML 235
QY 246 LNGRNSMESEVPRPKKETRMKRYHEVVRDTLHLTKRRKDPPEYMLTFPGNAVIVASRDF 305
DB 236 FTGENSELETEMPHKEERWKRYEYVNGKLTGTCTKMLPRLTCLPFIQSAGYFVTSREY 295
QY 306 VQHVLKPKSQOQLIEWYKDTYSPDEHLMATLQARAKMPSVPRNHPKDISMTSIRLVK 365
DB 296 VQVLENKIKQKIMEWAGQDYSPEYIWTATQIPREVPSIPASHKYDLSDMGAVARFVK 355
QY 366 WQGHESGIDIGAPYAPCSGIHQRAICVYGAGDLNMMLONHHLANKRPKVDNALQCLE 425
DB 356 WQYFEGDVSKGAPYPPCSGVHVRSVCLFGAGDLNMMLRKHHLFANKTIDVDVLAIGQCLDE 415
QY 426 YLRKYAI 433
DB 416 EHLRKAI 423

RESULT 12

AAB30297
ID AAB30297 standard; Protein; 427 AA.

AC AAB30297;

DT 12-FEB-2001 (first entry)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 08:15:46 ; Search time 19.87 Seconds

(without alignments)
538.420 Million cell updates/sec

Title: US-09-874-390-2

Perfect score: 2360
Sequence: 1 MVQMKRLCOLHYLMALGCMY.....NALQCLEXYLRKAYGTSL 438

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2360	100.0	438	US-09-233-506-2	Sequence 2, Appli
2	1142.5	48.4	428	US-07-955-041-4	Sequence 4, Appli
3	1142.5	48.4	428	US-08-227-455-4	Sequence 4, Appli
4	1142.5	48.4	428	US-08-472-482-4	Sequence 4, Appli
5	1142.5	48.4	428	US-08-487-069-4	Sequence 4, Appli
6	1142.5	48.4	428	US-09-233-506-3	Sequence 3, Appli
7	784	33.2	400	US-08-118-906-14	Sequence 14, Appli
8	784	33.2	400	US-08-486-196-14	Sequence 14, Appli
9	784	33.2	400	US-08-488-135-14	Sequence 14, Appli
10	784	33.2	400	US-08-474-065-14	Sequence 14, Appli
11	784	33.2	400	US-09-233-506-4	Sequence 4, Appli
12	447	18.9	126	US-08-118-906-4	Sequence 4, Appli
13	447	18.9	126	US-08-486-196-4	Sequence 4, Appli
14	447	18.9	126	US-08-488-135-4	Sequence 4, Appli
15	447	18.9	126	US-08-474-065-4	Sequence 4, Appli
16	426	18.1	126	US-08-118-906-2	Sequence 2, Appli
17	426	18.1	126	US-08-486-196-2	Sequence 2, Appli
18	426	18.1	126	US-08-488-135-2	Sequence 2, Appli
19	426	18.1	126	US-08-474-065-2	Sequence 2, Appli
20	278	11.8	64	US-09-233-506-10	Sequence 10, Appli
21	203.5	8.6	49	US-09-233-506-14	Sequence 14, Appli
22	126	5.3	33	US-08-118-906-8	Sequence 8, Appli
23	126	5.3	33	US-08-486-196-8	Sequence 8, Appli
24	126	5.3	33	US-08-488-135-8	Sequence 8, Appli
25	126	5.3	33	US-08-474-065-8	Sequence 8, Appli
26	118	5.0	33	US-08-118-906-6	Sequence 6, Appli
27	118	5.0	33	US-08-486-196-6	Sequence 6, Appli

28	118	5.0	33	1	US-08-488-135-6	Sequence 6, Appli
29	118	5.0	33	2	US-08-474-065-6	Sequence 6, Appli
30	109	4.6	22	1	US-08-118-906-12	Sequence 12, Appli
31	109	4.6	22	1	US-08-486-196-12	Sequence 12, Appli
32	109	4.6	22	1	US-08-488-135-12	Sequence 12, Appli
33	109	4.6	22	2	US-08-474-065-12	Sequence 12, Appli
34	103	4.4	316	1	US-08-597-226-12	Sequence 12, Appli
35	103	4.4	316	1	US-08-746-682A-12	Sequence 12, Appli
36	101	4.3	788	2	US-08-907-166-6	Sequence 6, Appli
37	100.5	4.3	2555	3	US-09-058-489-36	Sequence 36, Appli
38	94.5	4.0	253	3	US-08-929-329-9	Sequence 9, Appli
39	94.5	4.0	619	4	US-09-153-310-42	Sequence 42, Appli
40	92.5	3.9	2289	3	US-09-051-019-2	Sequence 2, Appli
41	91	3.9	3135	1	US-08-323-170B-2	Sequence 2, Appli
42	91	3.9	3135	4	US-08-954-441-2	Sequence 2, Appli
43	90.5	3.8	616	4	US-08-965-902A-2	Sequence 2, Appli
44	90	3.8	1128	1	US-08-111-939-2	Sequence 2, Appli
45	90	3.8	1194	4	US-09-092-508-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1	US-09-233-506-2	Sequence 2, Application US/09233506
Patent No. 6136580	GENERAL INFORMATION:	
APPLICANT: Fukuda, Minoru		
APPLICANT: Yeh, Jium-Chern		
TITLE OF INVENTION: A Beta-1-6-N-Acetylglucosaminyltransferase That Forms		
FILE REFERENCE: P-1/3 3415		
CURRENT APPLICATION NUMBER: US/09/233, 506		
CURRENT FILING DATE: 1999-01-19		
NUMBER OF SEQ ID NOS: 14		
SOFTWARE: PatentIn Ver. 2.0		
SEQ ID NO 2		
LENGTH: 438		
TYPE: PRT		
ORGANISM: Homo sapiens		
US-09-233-506-2		
Query Match	100.0%	Score 2360; DB 4; Length 438;
Best Local Similarity	100.0%	Pred. No. 6.9e-230;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY 1 MVQMKRLCOLHYLMALGCMYMLATVALKLSFRLKCDSDHGLGLESRESOSQYCRNLYNFL 60		
DB 1 MVQMKRLCOLHYLMALGCMYMLATVALKLSFRLKCDSDHGLGLESRESOSQYCRNLYNFL 60		
QY 61 KLPKRSINSGVTRGDOEAVLQALNNLEVKKKREPTDTHYSLTFDCEHFAERKFI 120		
DB 61 KLPKRSINSGVTRGDOEAVLQALNNLEVKKKREPTDTHYSLTFDCEHFAERKFI 120		
QY 121 QPLSKVEVERPIAYSMYIHEKIEFERLRAVAPQNYCVHNDKSPETFEKAVKAI 180		
DB 121 QPLSKVEVERPIAYSMYIHEKIEFERLRAVAPQNYCVHNDKSPETFEKAVKAI 180		
QY 181 SCFPVVFASLKLVRVYASWSRVQADLNCMEDLLQSSVPWKYFTLNTCGTDPPIKSNABV 240		
DB 181 SCFPVVFASLKLVRVYASWSRVQADLNCMEDLLQSSVPWKYFTLNTCGTDPPIKSNABV 240		
QY 241 QALMKLNGRNMESVPPKHKETRWKYHFEVVRDTLHTNNKKDDPPYNTLMTFGNAVY 300		
DB 241 QALMKLNGRNMESVPPKHKETRWKYHFEVVRDTLHTNNKKDDPPYNTLMTFGNAVY 300		
QY 301 ASRPVQVHVLNPKSQOILEVVKDTYSPDEHLMATLQARMPGSPVHPKYDISDMTSI 360		
DB 301 ASRPVQVHVLNPKSQOILEVVKDTYSPDEHLMATLQARMPGSPVHPKYDISDMTSI 360		
QY 361 ARLVKQGHGDDIKGAPYAPDCGIHQRAICVYGAGDLNMMLONHHLANKFDPKVDNDA 420		

```

Db      361  AATLVKMGQHEGDIKDGPAPYPCSGIHQRALICVYGAGDJLNMMLQNHHLLANKEDPKYDDNA 420
Oy      421  LACLEXYLRKYKAITYTEL 438
        |||||||
Db      421  LACLEXYLRKYKAITYTEL 438

RESULT 2
US-07-955-041-4
; Sequence 4, Application US/07955041
; Patent No. 5360733
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, MINORU
; APPLICANT: BIERHUIZEN, MARTI FA
; TITLE OF INVENTION: A NOVEL BETAI-6
; TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE
; TITLE OF INVENTION: LEDROSITALIN AND A METHOD FOR CLONING PROTEINS HAVING
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,041
; FILING DATE: 19921001
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHERIN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9294
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 428 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-955-041-4

```

Query Match	48.48%	Score 1142.5	DB 1	Length 428
Best Local Similarity	57.58%	Pred No. 6.3e-107		
Matches	211	Conservative	57	Mismatches 98
			Indels	1
			Gaps	1

QY	68	INCGSVTRGDDQAVLQALINLNLEVK-KKREPPTDHYLSLTRDCHEFKAKRRFTQDPLSK	126
		: : : : : : : : : : : : : : :	
	DB	: : : : : : : : : : : : : : :	
	57	INCTFYLOQDVAEIKVKKEILITVFKKRRPMTPDYINMTSCSSFFIRKRYIVELPK	116
QY	127	EEVEPPFASVNIHEKIEKFERLLRVVAPQNIYCVHVDKSEPTKEAVKRIISCFPNV	186
		: : : : : : : : : :	
DB	117	EEAEPPFIASIVYHHKIEHMDRLRLRIYTPQNFYCVHVTKSEDTSLAVMGIAGSFSNV	176
QY	187	FLASLKLVRVYVASWSRVCQADLNCMEDLIQSSVPWKYELNTGCTDPEIKSNAEQALKML	246
		: : : : : : : : :	
DB	177	FVASRLESVYVASWSRVCQADLNCMDKDLVAMSNWKLIMCGMDPEIKTNLEIVRKLKLL	236
QY	247	NGRNSMSESVPPKHKETPRMKYHFEVAVROTLLHLLTNKKKDDPPNLTTFGTGNATVARSDEV	306
		: : : : : : : : :	
DB	237	MGENNLETFRRMSHKEERKKRRRYEVNGLRTGTGYKMLPRLDTLFFSSAFAVYSREYV	296
QY	307	QHYLKNPKSQQLIENWKDTYSPDEHLMATLQARRMPGSPVPHNPKYDIDSMTSLARLVKM	366

Db	297	GVLINNEKIQKIMWAQDTYSPDELYMTIQRIPEVPGSLPASHKHYDLSDMQAVAFVWK	356
OY	367	QGHEDDIDKGPARYPCSSITHORALCVGAGGLNMLNMLNHLANKEPKPYDDNALQCLDE	423
Db	357	QYFEDDSKSGAPYPCCDGVHRSVCIFGAGDNLMLLRNHLANKFPYDVDLFAIQCLDE	416
OY	427	YLRYKAI	433
Db	417	HLRHKAL	423

RESULT 3
 US-08-227-455-4
 : Sequence 4, Application US/08227455
 : Patent No. 5624832
 : GENERAL INFORMATION:
 : APPLICANT: FUKUDA, MINORU
 : APPLICANT: BIERHUIZEN, MARTI FA
 : TITLE OF INVENTION: A NOVEL BETAL-6
 : TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE
 : TITLE OF INVENTION: LEUKOSTALIN AND A METHOD FOR CLONING PROTEINS HAVING
 : ENZYMATIC ACTIVITY
 : NUMBER OF SEQUENCES: 8
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: CAMPBELL AND FLORES
 : STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
 : CITY: SAN DIEGO
 : STATE: CALIFORNIA
 : COUNTRY: USA
 : ZIP: 92122
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/227,455
 : FILING DATE: 14-APR-1994
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: CAMPBELL, CATHERYN
 : REGISTRATION NUMBER: 31,815
 : REFERENCE/DOCKET NUMBER: P-LJ 9957
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 619-535-9001
 : TELEFAX: 619-535-8949
 : INFORMATION FOR SEQ ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 428 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-227-455-4

	Query Match	48.4%	Score 1142.5;	DB 1;	Length 426;	
	Best Local Similarity	57.5%;	Pred No. 6.3e-107;			
	Matches 211; Conservative	57;	Mismatches 98;	Indels 1;	Gaps 1	
Oy	68	INCGVTRGDDEAVLQALINLNLEVK-KRREPPTDTHYLSLRDCDEHFKAERFETOPPLSK	126			
		: :	:	:		
Dd	57	INCFVTLGVADVNEIQKVKLEITLVFKRPKPDPDDYNMTSDCSSFIIRRRKYIVELPKL	116			
Oy	127	EEVEEPIIASWVIHKKIENFERLLAAVVAPOVICYHNDEKSPEFFKEAKVAKIIISOEPNV	186			
		: :				
Dd	117	EAEAPPIASIVAHKIKIEDRLRIATYPOMNFCVHNVDKESDYSLAAMVGASFSENV	176			
Oy	187	FIAKLVLVRVVASMSRGVADNLNCMEDIALOSSVPARKYFLTCTGDFPIKSNMAEQVALTKML	246			
		:				
Dd	177	FVASHLESVVVASWSRGVQADLNCMDLYAMSANWKRYLINLCMPDPPIKTNLEIVRKLL	236			
Oy	247	NGRNMESEVPPKHKETRKTYHEEVAVRDTLHLTNKKKDPDPPIINLMFTGNAYIVASRDV	306			

Db 237 MGENNLETERMPSHKEERKRYEVANGKLTNTGTVMKLPLETPLESGSAFVVSREV 296
Qy 307 QHVLNPKSQOLIEVVKDTYSPDEHLMTLQARMMPGSVPHRPKYDIDSMTIARLYKM 366
Db 297 GVLQNEKIQKLEMAOQTYSPDEYLMATIQRIPEVPSLPASHRYDLSDMQAVARFYKM 356
Qy 367 QGHEGDIDKAPYAPPCSGIHORALCVGAGDLNMMLOHHLANKFDPKVDNALQCLEE 426
Db 357 QYFEGDVSKGAPYPCDGVHVRSCIFGAGDLNMMLRKHHLFANKFDVDVDFALQCLDE 416
Qy 427 YLRKAI 433
Db 417 HLRKAL 423

RESULT 4
US-08-472-482-4
; Sequence 4, Application US/08472482
; Patent No. 5658778
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, MINORU
; APPLICANT: BIERHUIZEN, MARTI FA
; TITLE OF INVENTION: A NOVEL BETAL-6
; TITLE OF INVENTION: N-ACETYLDUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
; TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
; TITLE OF INVENTION: ENZYMACTIC ACTIVITY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,482
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/955,041
; FILING DATE: 01-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHERYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9294
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 428 amino acids
; TYPE: amino acid.
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-472-482-4

Query Match 48.4%; Score 1142.5; DB 1; Length 428;
Best Local Similarity 57.5%; Pred. No. 6.3e-107;
Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;

Qy 68 INCSGVTGDOEAVLQALINLEVA-KRREPPTDTHYSLTRDCHEFAERKFFPLSK 126
Db 57 INCTVVLQGVNLEIKVLEILITVFKRRPRTPDDYINMTSDCSFTRKRYIVEPLSK 116
Qy 127 EEVEEPPIAYSWIHEKINFERLLRAVAPQNIYCVHVDKSPERFKKAVKAIISCFNV 186
Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;

Db 117 EEAEPPIAYSIYVHHKIEMLDLRLAITYPQNFVCHVDTKESDYLAVMGIASCFSNV 176
Qy 187 FIASKLVRVYVASMGRVADLNCMEDLQSSVPMKYFLNTCTGDPPIKSMAEMOALKM 246
Db 177 FVASHLESYVYASMSRVQADLNCMDLYAMSNMRYLLNLGMDPIKTNLEIVRKILL 236
Qy 247 NGRNMESEVPPKHEETRKXHEEVVRODTLHLTNKKKODPPYNNLTFTGNAYIVASRDFV 306
Db 237 MGENNLETERMPSHKEERKRYEVANGKLTNTGTVMKLPLETPLESGSAFVVSREV 296
Qy 307 QHVLNPKSQOLIEVVKDTYSPDEHLMTLQARMMPGSVPHRPKYDIDSMTIARLYKM 366
Db 297 GVLQNEKIQKLEMAOQTYSPDEYLMATIQRIPEVPSLPASHRYDLSDMQAVARFYKM 356
Qy 367 QGHEGDIDKAPYAPPCSGIHORALCVGAGDLNMMLOHHLANKFDPKVDNALQCLEE 426
Db 357 QYFEGDVSKGAPYPCDGVHVRSCIFGAGDLNMMLRKHHLFANKFDVDVDFALQCLDE 416
Qy 427 YLRKAI 433
Db 417 HLRKAL 423

RESULT 5
US-08-487-069-4
; Sequence 4, Application US/08487069
; Patent No. 5684134
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, MINORU
; APPLICANT: BIERHUIZEN, MARTI FA
; TITLE OF INVENTION: A NOVEL BETAL-6
; TITLE OF INVENTION: N-ACETYLDUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
; TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
; TITLE OF INVENTION: ENZYMACTIC ACTIVITY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,069
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/955,041
; FILING DATE: 01-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHERYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9294
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 428 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-069-4

Query Match 48.4%; Score 1142.5; DB 1; Length 428;
Best Local Similarity 57.5%; Pred. No. 6.3e-107;
Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;

```

QY 68 INSCVTGCGQEAVALIINNLEVK-KKREPFDTHYLSLTRDCEHFKAEKRTIOPLSK 126
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 57 INCTKVLGGVNEIOKKEILTYKFKRRPRPDDYINMTSDCSSFKRKRKYIVEPLSK 116
QY 127 EEEFPPAYSAMVHEKJENEFERLLRAYAPONICYVHDEKSPETFEAVKAIISCPNV 186
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 117 EEEFPPAYSAMVHEKJENEFERLLRAYAPONICYVHDEKSPETFEAVKAIISCPNV 176
QY 187 FIASKLVRVYASMSRVQADLNCMEDLLQSSVPMKYFLNTCGTDFPIKSAEWOALKML 246
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 177 FVASRLSESVYASMSRVQADLNCMDLYAMSAWKYILINCGMDFPIKTLNLEIYRKAKL 236
QY 247 NGNSMSSEVPKPKHETRMKYHEEVVADTLHLNKKKDPPEYMLTMTGNAYIVASRDEV 306
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 237 MGENNLETETMPSPHKEERMKRYEVNGKLTNGTYKMLPPLTETPLFSGSAFYVSREYV 296
QY 307 QHVLKNPKSOQLLEWYDTSPEHLMATLQRAKMPGSPVNHKPYISMTSIAKLVK 366
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 297 GYVLQNKIKOKLEWADDTSPDEYLMATLQRIPEVPGSLPASHKIDLSMOAVARVVK 356
QY 367 QGHEGIDKAGAPYAPCSGIIHQRAICVYGAGDLNMMLONHHLANKRPDYDNLQCLEE 426
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 357 QYEGGVSKGAPYPCGIVHVSVCIFGAGDLNMMLRKHHLPANKPVDVDFALQCLDE 416
QY 427 YLRKAI 433
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 417 HLRKAL 423

```

RESULT 6

```

US-09-233-506-3
; Sequence 3, Application US/09233506
; Patent No. 6136580
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; TITLE OF INVENTION: A Beta-1-6-N-Acetylglucosaminyltransferase That Forms
; FILE REFERENCE: P-LJ 3415
; CURRENT APPLICATION NUMBER: US/09/233,506
; CURRENT FILING DATE: 1999-01-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 428
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-233-506-3

```

Query Match 48.4%; Score 1142.5; DB 4; Length 428;
 Best Local Similarity 57.5%; Pred. No. 6.3e-107;
 Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;

```

QY 68 INSCVTGCGQEAVALIINNLEVK-KKREPFDTHYLSLTRDCEHFKAEKRTIOPLSK 126
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 57 INCTKVLGGVNEIOKKEILTYKFKRRPRPDDYINMTSDCSSFKRKRKYIVEPLSK 116
QY 127 EEEFPPAYSAMVHEKJENEFERLLRAYAPONICYVHDEKSPETFEAVKAIISCPNV 186
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 117 EEEFPPAYSAMVHEKJENEFERLLRAYAPONICYVHDEKSPETFEAVKAIISCPNV 176
QY 187 FIASKLVRVYASMSRVQADLNCMEDLLQSSVPMKYFLNTCGTDFPIKSAEWOALKML 246
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 177 FVASRLSESVYASMSRVQADLNCMDLYAMSAWKYILINCGMDFPIKTLNLEIYRKAKL 236
QY 247 NGNSMSSEVPKPKHETRMKYHEEVVADTLHLNKKKDPPEYMLTMTGNAYIVASRDEV 306
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 237 MGENNLETETMPSPHKEERMKRYEVNGKLTNGTYKMLPPLTETPLFSGSAFYVSREYV 296
QY 307 QHVLKNPKSOQLLEWYDTSPEHLMATLQRAKMPGSPVNHKPYISMTSIAKLVK 366
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

```

```

Db 297 GYVLQNKIKOKLEWADDTSPDEYLMATLQRIPEVPGSLPASHKIDLSMOAVARVVK 356
QY 367 QGHEGIDKAGAPYAPCSGIIHQRAICVYGAGDLNMMLONHHLANKRPDYDNLQCLEE 426
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 357 QYEGGVSKGAPYPCGIVHVSVCIFGAGDLNMMLRKHHLPANKPVDVDFALQCLDE 416
QY 427 YLRKAI 433
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 417 HLRKAL 423

```

RESULT 7

```

US-08-118-906-14
; Sequence 14, Application US/08118906
; Patent No. 5484590
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Hierulzen, Marc F.A.
; TITLE OF INVENTION: Expression of the Developmental I
; TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,906
; FILING DATE: 09-SEP-1993
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9526
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 335-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-118-906-14

```

Query Match 33.2%; Score 784; DB 1; Length 400;
 Best Local Similarity 42.3%; Pred. No. 9.1e-71;
 Matches 163; Conservative 62; Mismatches 126; Indels 34; Gaps 9;

```

QY 56 LYNFLKLPARKSINSGVTGCGQEAVALIINNLEVKKKREPFDTHYLSLTRDCEHFK 115
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 24 VFNFGDPSFORLINDPLRLQ--VCTSRING-----KTFELKKNKLMNHEKSCKEYL 77
QY 116 ERKEIOPLSKKEVEERPIAYSAMVHEKJENEFERLLRAYAPONICYVHDEKSPETFEK 175
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 78 QSHYITAPLSKEADRPPLATIVIHNFDFRALFRATVMPONICYVHDEKATTEKKA 137
QY 176 VKAIISCPNVTASKLVRVYASMSRVQADLNCMEDLLQSSVPMKYFLNTCGTDFPIK 235
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 138 VEOILSCFPNAPFLASKMEPVYIGISRLQADLNCIDLSAFESWKKYVINTCGODPELKT 197
QY 236 NAEWOALKMLNCRNMSSESVPKPKHETRMKY-HFE-----YVADTLHLNKKKDDP 286
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 198 NKEIYOLKGFKGKNTTPGVLPRAHAIGRKTVYHDELKELSYLR-----TTLAKPP 252

```

Qy	287	PYNTLMFGNVIYAASRPFVOHVLKPNPKSQOOLIEWVKDTPSPDELTMTLOARMMPSV	346
Db	253	PHNLITIGSAIVALSREFANFVLHPRAVDLLQSKDTFSDFEHFWTLNRFIPGVPGSM	312
Qy	347	PNHPRKYDSIDMTSIARLVKMOGHGEGDDIKGAPYAPCSGIHORAIICVGAGDLNMILONHH	406
Dy	313	PN-----ASWTGNLRAIKMSDME---DR---HGSGCHGYHYGCITCYNGDKMLVNSPS	360
Qy	407	LANKFDPKYDDNALOCLEIYLRK	431
Dy	361	LFAKFELNYPLVECLE-LRRHR	383
RESULT 8			
US-08-486-196-14			
Sequence 14, Application US/08486196			
Patent No. 5731420			
GENERAL INFORMATION:			
APPLICANT: Fukuda, Minoru			
APPLICANT: Blerhuken, Marti F.A.			
TITLE OF INVENTION: Expression of the Developmental I			
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a			
TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family			
NUMBER OF SEQUENCES: 14			
CORRESPONDENCE ADDRESSES:			
ADDRESSEE: Campbell and Flores			
STREET: 4370 La Jolla Village Drive, Suite 700			
CITY: San Diego			
STATE: California			
COUNTRY: USA			
ZIP: 92122			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Patentln Release #1.0, Version #1.25			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/486,196			
FILING DATE:			
CLASSIFICATION: 424			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/118,906			
FILING DATE: 09-Sep-1993			
ATTORNEY/AGENT INFORMATION:			
NAME: Campbell, Cathryn A.			
REGISTRATION NUMBER: 31,815			
REFERENCE/DOCKET NUMBER: P-LJ 9526			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: (619) 535-9001			
TELEFAX: (619) 535-8949			
INFORMATION FOR SEQ ID NO: 14:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 400 amino acids			
TYPE: amino acid			
TOPOLOGY: linear			
MOLECULE TYPE: protein			
US-08-486-196-14			
Qy	Query Match	33.2%; Score 784; DB 1; Length 400;	
Dy	Best Local Similarity	42.3%; Pred. No. 9,le-71;	
Matches	163; Conservative	62; Mismatches 126; Indels 34; Gaps 9;	
Qy	56 LYNFLKLPKAKSINSQGTBGDAEAVLQALINNLEVKKKRREFFDTHTLSLRDCEHFKA	115	
Dy	24 VFNRGDPSPFORLNISDLRLTQ--VCISFTING----KTRELWNKKMLTHKSSSEKEYLT	77	
Qy	116 ERKEIOPLSLKEEEVFPIAYSVMVHEKIENEFRLRAYAYAPONICYCVNDEKSPTFEKA	175	
Dy	78 OSHVITAPLSKEADDFLAYIMVLIHHNDTFARLFRAIYMPONICYCVNDEKATTEFKDA	137	
Qy	176 VKAIIISCPPNVFIASKLVLRVVYASKRVAODLNCMEDILLQSSVPKKFYLATNGCTDPIKS	235	

[illegible]

NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell and Flores
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/488,135
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/118,906
 FILING DATE: 09-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LJ 9526
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 126 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 JS-08-488-135-4

```

Only Match 18.9%: Score 447; DB 1; Length 126;
Best Local Similarity 65.9%: Prod. No. 1.5e-37;
Matches 83; Conservative 15; Mismatches 28; Indels 0; Gaps 0.

QY 119 FIOPLSKSEEEFFPIASVMVIEHKIENFERLRLRAYAPQNICVHVDKSPETFEAKVA 178
      1 YIYELPSNSEEAEFFPIASVIVVHHKITEMDLRLRALIYMQNEFCVHVDRKSEDSYLAAVNG 60
      :| ||||| |||||:::| || :|||::| || ||||| ||::: ||
QY 179 IISCPNVEFIASKLRRVYVYASRSVQADLNCMEDLLQSSVPMKYFLNCTGIDFPIKSNAE 238
      61 IASCSNFVEFASRLSESVYASRSVQADLNCMKRDLYAMSANKRYLINCGMDFPIKTNDE 120
      || |||||::| ||||| |||||::| || ||::| |||||::|
QY 239 MVQALK 244
      ||::||
Db 121 IYRKIK 126

RESULT 15
US-08-474-065-4
: Sequence 4, Application US/08474065
: Patent No. 5830465
: GENERAL INFORMATION:
: APPLICANT: Bierhuizen, Marti F.A.
: APPLICANT: Bierhuizen, Marti F.A.
: TITLE OF INVENTION: Expression of the Developmental I
: TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
: TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell and Flores
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,065
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-IJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-474-065-4

```

	Query Match	18.9%;	Score 447;	DB 2;	Length 126;	
	Best Local Similarity	65.9%;	Pred. No. 1.5e-37;			
	Matches	83;	Conservative 15;	Mismatches 28;	Indels 0;	Gaps 0.
QY	119	FIOPLSKEEVEFFPIAYSMVIEHKIENFERLRAVAYADONIVCYHVDKSPETFEKAANYA	178			
		: : : : : : :				
Dd	1	YIVELSLKEEAEFFPIAYSIVVHHKTEMLDRLLRALTYMPQNFCVHYDFKSEDSYLAAVNG	60			
QY	179	IISCFPNVFIASKLVRYVVYASRSRVQADLNCKMEDILQSSVPKKYFLNLTGCTFPPIKSNAE	238			
		: : : : : :				
Dd	61	IASCNSNFVASRLESVVYASRSRVQADLNCKMRDLYAMSANKKYLINLCGMDFPIKTNLE	120			
QY	239	MVOALX 244				
		:				
Dd	121	IVRKLR 126				

Search completed: September 26, 2002, 09:45:39
Job time: 5393 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 08:19:26 ; Search time 38.79 Seconds
(without alignments)
1084.999 Million cell updates/sec

Title: US-09-874-390-2

Perfect score: 2360
Sequence: 1 MYQWKRICQLHYLMALGCMY.....NALQCLEYLRKAYGTSL 438

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1142.5	48.4	428	2 A46293	beta-1,3-galactosyl
2	784	33.2	400	2 A46297	beta-1,6-N-acetylgl
3	379	16.1	472	2 T22188	hypothetical prote
4	371	15.7	445	2 T21262	hypothetical prote
5	367	15.6	401	2 T24929	hypothetical prote
6	352	14.9	425	2 T20207	hypothetical prote
7	349.5	14.8	402	2 T24930	hypothetical prote
8	344.5	14.6	454	2 T21261	hypothetical prote
9	271	11.5	470	2 T32137	hypothetical prote
10	267	11.3	467	2 T21408	hypothetical prote
11	267	11.3	483	2 T89364	hypothetical prote
12	259.5	11.0	322	2 D87603	glycosyl transfera
13	254	10.8	454	2 T25382	hypothetical prote
14	248.5	10.5	513	2 T20068	hypothetical prote
15	232	9.8	489	2 T24013	hypothetical prote
16	219	9.3	459	2 T24742	hypothetical prote
17	213.5	9.0	486	2 T21566	hypothetical prote
18	212.5	9.0	367	2 T28892	hypothetical prote
19	210.5	8.9	478	2 T27714	hypothetical prote
20	205.5	8.7	753	2 T24745	hypothetical prote
21	199	8.4	472	2 T33384	hypothetical prote
22	177	7.5	406	2 T86571	hypothetical prote
23	176	7.5	448	2 D85042	probable glycosyla
24	166	7.0	395	2 B86735	unknown protein F2
25	164.5	7.0	447	2 T00906	hypothetical prote
26	157	6.7	434	2 T51450	hypothetical prote
27	150	6.4	384	2 T08940	hypothetical prote
28	145	6.1	371	2 T21819	hypothetical prote
29	125	5.3	630	2 T02524	probable RING zinc

30	116	4.9	298	2 AD2346	hypothetical prote
31	108	4.6	1201	2 H69898	hypothetical prote
32	105	4.4	787	2 C75068	probable beta-gala
33	104.5	4.4	300	1 ND8P7	exodeoxyribonuclea
34	101.5	4.3	302	1 S07503	exodeoxyribonuclea
35	100	4.2	897	2 G89923	hypothetical prote
36	99.5	4.2	876	2 G89923	DNA polymerase I
37	99.5	4.2	946	2 T31488	hypothetical prote
38	99	4.2	690	2 C96572	protein F12M16.4
39	97	4.1	823	2 C83905	hypothetical prote
40	97	4.1	987	1 T0B530	transposase - Baci
41	96.5	4.1	779	1 S61571	suppressor protein
42	95.5	4.0	735	2 A82817	pentaphosphate qua
43	95	4.0	932	2 T25008	hypothetical prote
44	94.5	4.0	498	2 T47919	hypothetical prote
45	94.5	4.0	671	2 S61693	probable membrane

ALIGNMENTS

RESULT 1

A46293 beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase

C:Species: Homo sapiens (man)

C:Date: 20-Oct-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000

C:Accession: A46293

R:Bierhuizen, M.F., Fukuda, M.

Proc. Natl. Acad. Sci. U.S.A. 89, 9326-9330, 1992

A:Title: Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcN

A:Reference number: A46293; MUID:93028457

A:Accession: A46293

A:Molecule type: mRNA

A:Residues: 1-428 <BIE>

A:Cross-references: GB:M97347; NID:g183440; PIDN:AAA35919.1; PID:g183441

A:Experimental source: HL-60 cells

A:Note: sequence extracted from NCBI Backbone (NCBIW:115900, NCBIIP:115901)

C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match	Best Local Similarity	Score	DB 2;	Length	428;
Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;					
QY 68	INCSGVTGGDQAVLQATLNNLEVK-KKREPTDHHVSLTRDCSHFAERKFIQPIK 126				
DB 57	INCTVLOGDVNEIOKVKLEILTVKFKRPRTDPDYLMTSDCSSFKRKRYVEPLSK 116				
QY 127	EEVEPPIAVSWVHHEIKENFERLRAVYAPQNIYCVHDEKSPETFEKAVKAIIISCFPNV 186				
DB 117	EEAEPPAVSIYVHHKIEMLDRLAIRATMPQFVCHVDTSKEDSYLAAVMGIASCFSNV 176				
QY 187	FIASKLVRVYASWSRVQADLNCMEDLLOSSVPKWFYFLNTGCTDPPIKSNAEMVALKML 246				
DB 177	FVASKLVRVYASWSRVQADLNCMEDLLOSSVPKWFYFLNTGCTDPPIKSNAEMVALKML 236				
QY 247	NGRNSMESEVPKHKETEMKRYHEFVVRDTLHNTNKKDPPYNLMPFGNAYTVASRQFV 306				
DB 237	MCENNLETERMSHKEEMKRYHEFVVRDTLHNTNKKDPPYNLMPFGNAYTVASRQFV 296				
QY 307	QHVLTNPKRSQOLIEVWKDTPSPDEHLMATLQARAWMPGSPVHPKDYDISMTSIRLWK 366				
DB 297	GVLVONEKIOKIMENAQDTYSDEFLMTIQRIPEVPSLASHKYDLSDMQAVARFVKW 356				
QY 367	QGHESDIDKGAAPACSSIHORATCVGAGDLNMLQNHHLANKFDPKVDNALQCLEE 426				
DB 357	QYFEEDVSGKAPYPCDGVHVSVCIFGAGDLNMLRKHHLFANKFDVVDLFAIQCIDE 416				
QY 427	YLRYKAI 433				
DB 417	HLRRKAL 423				

Query Match	33.28	Score	784	DB 2:	length	400;			
Best Local Similarity	42.38	Pred.	No. 1.2e-52;						
Matches	163;	Conservative	62;	Mismatches	126;	Indels	34;	Gaps	9

RESULT 3
T2218

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22188
R:Colles, L.
submitted to the EMBL Data Library, September 1994
A:Reference number: Z19528
A:Accession: T22188
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-472 <RTL>
A:Cross-references: EMBL:Z37092, PIDD:CAA85457.1, GSPDB:GN00020, CESP:F44F4.6
A:Experimental source: clone F44F4
C:Genetics:
A:Gene: CESP:F44F4.6
A:Map position: 2
A:Introns: 57/2; 105/2; 185/3; 261/3; 319/1; 370/1; 412/3

```

0Y 31 FRKCDSPHLESE-----SOSYCRHILNFKL-----PAKS-----INSGV 73
Db 13 FRKFCCLQGLMNVRRKRIISQFISSTPIYIVRLGFDVSLVPEKKLENTPLEGNSV 72
QY 74 TRGDQEAVALQALINLNEVKKKREPFDTHY-----LSLTRDCEHFKABKKFIQPLSK 126
Db 73 INGE-----NRRKISRARQ--WDMHEFMEVHEHIFNSNNVSTIDKFYFNFRIPSSR 122
QY 127 EEVEERPLASVNIHEKLENFERRLARAYARONICVAVDEKSPTEKAVKAIISCPN- 165
Db 123 EEAEPLRPLAVGLVYKVTIYOVLQMSLEFYQDPMHCITVDQOSPNEYKSVIQLSPCEPN 182
QY 186 -VFIAKLVRVYVSWSVQADLNCMEDLLQSSVPKRYFLNTGCTDPIKSNAMVQALK 244
Db 183 HVFIGEP---SQMSFGILKAVYICFPMWLSKSKKMKRYIYOLSTQDLPITNLEWRIK 239
QY 245 MLCNGRSMESFVPRKHETRMKYHEFVYVDTLHLTNKKKODPRYVNLTMFTGNAYIASRD 304
Db 240 ALNG--SMTGYVST-----FEYDR--YKNMEGVLPR--MFVYKSSSVVPR 281
QY 305 FVQVYHLKPKKSOQLEWYKDTYSPDEHMLATIQ-----RAR---WMDGSPVNH 349
Db 282 GADYLISSPRQOKLKLKLTWIPDESWFVJLGSPLMLPVGSIYKRDILMLKKNKLR 341
QY 350 PKYDISDNTIA-----RLVKWQGHGIDIKGAPYAPCSGIHQRAICVYAGADLNMWL 402
Db 342 PRYE--NFVNSIGTGYIGRYQWGMQ-----KECFGAKKDCSVCFVEDIEIM 388
QY 403 QNHHLANKRPDKVDNNAOLCELEUYRYKAI 433
Db 389 TRPELVANKLYLEQRPAPFOMLEEVNRRSL 419

```

```

RESULT 4
T21262
hypothetical protein F22D6.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21262
R:Wilkinson, J.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19397
A:Accession: T21262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-445 <MIL>
A:Cross-references: EMBL:Z71262; PIDN:CAA95817.1; CSPDB:GND00019; CSPS:F22D6.11
A:Experimental source: clone F22D6
C:Genetics:
A:Map position: 1
A:Gene: CSPS:F22D6.11
A:Introns: 65/3; 95/2; 136/3; 180/3; 256/3; 281/3; 403/3

```

Query Match	15.7%	Score 371;	DB 2;	Length 445;
Best Local Similarity	25.5%	Pred. No. 8.5e-21;		
Matches	98;	Conservative	76;	Mismatches 149; Indels 62; Gaps 9
Qy	68	INCSVTVTGDDGAEVALQALINLIEVKKKRPFFDTRHLSLITRCEHFKAARKTQIPLSLKE	127	
Db	62	INCLNFIENGSKRRNDRLOTRINSRISDK-----LLVSTDRCCQLTKSLFRKNPKLSPSE	113	
Qy	128	EEVEFPIAYSMVITHEK-----IENFERLLRAYVAPONITYCVAHUDEKSPETFEKAVKAIISIC	182	
Db	114	EESEFPLSLGGLVYKKRLSGVTLQIRVLFEMLSIYHPQNETGIAANGENSAPRIFQULNLRKLSNC	173	
Qy	183	FPNWFIAASKLVRYVYASMSRVOADLNCMEDLLQSSVPMKYFLFNTGCTDFPISNAEMVQA	242	

Db 174 FSNHIF-MKRPPIIDMGSHIINSAYDCLIEFLSHLSDMRFFOYLSGVDIPLKTNLEWVOI 232
Qy 243 LKMLNGSNSESEVPPKHKETRMKYHEVVRDTHLTKKKDPPRYMLTJMTGNATYVAS 302
Db 233 LKHLNGTANVEIK-PYQYRLRGR-----NEIOSPLP-----LFKSSLSLIIIP 274
Qy 303 RDEVOYHLKPNKRSQOLIEWKDYSPDEHLMATLORARMMGVSYPNHPKIDISDTSIAR 362
Db 275 REANHLSSSSIPQOLLEPLRNTGIDAEQFWGTL-----FGNKLFPDIPGSLN 322
Qy 363 LVKMOGHEGIDKCAPAP-----CSGHORAIQYAGADLMMMLQNH 405
Db 323 EKEMISYKNNEETMLTPYTDGMYRISRDQIWSKPNCHNYMKAGSCVFGLIDVPLRLSK 382
Qy 406 HLANKFDPKVDNALOC-LEEYLK 429
Db 383 ALVAHKFYLKSEPPAYFCLLEHRR 407

RESULT 5
T24929

hypothetical protein T15D6.2 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T24929

R/Dobson, R.
submitted to the EMBL Data Library, November 1996

A/Reference number: Z19956

A/Accession: T24929

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-401 <M12>

A/Cross-references: EMBL:283125; PTDN:CA05620.1; GSPDB:GN00019; CESP:T15D6.2

A/Experimental source: clone T15D6

C/Genetics:

A/Map position: 1

A/Introns: 45/3; 75/2; 111/3; 151/3; 227/3; 252/3; 296/2; 359/3

Query Match 15.6%; Score 367; DB 2; Length 401;
Best Local Similarity 25.5%; Pred. No. 1.3e-20;
Matches 97; Conservative 74; Mismatches 143; Indels 66; Gaps 8;

Qy 57 YNFKLPAKR-----SINCGVTRGDOEAVLQAILNMLEYKKRREPTDTHYLSLT 107
Db 22 YLFVAKSSRPPTDLLDDLEINTNIIQGLKNEKOLEINTKMTEDKIMNSTDR----- 75
Qy 108 RDCHEFAERKFIQPLSKEEVEPPRIASVNIHERKIEFERLIRAVYAPQNYCVHDEK 167
Db 76 --CHTVSMFRFNVPLSEEARPLSLGILVYKELSGVLFMLSSITQPOMEYCIAVGEN 133
Qy 168 SPEFFKAVKAIISCFFPNVFLASKLVAVYASMSRVOADLQSSVPMKYFLNTC 227
Db 134 SASFTLLLELSDCFPN-----KRPIWGSYEIINSVYDCLFKLSHKSMKYFYLS 188
Qy 228 GTDPIKSNMVOALKMLNGRSMSESVPRKHETRMKYHEVVRDTHLTKKKDPP 287
Db 189 GVDPLKTNLEWVAILSLNGTANVEIKYENRR-----LQONETESPLP 234
Qy 288 YNLTMTGNATYVASRDFVOHLKPNKRSQOLIEWKDYSPDEHLMATLQPARMM---PG 344
Db 235 ----LFKSSLSLIIIPKRAANYLASSSIPOQLLEFLRNTWVADEQFWGTLPGNKLFDVPG 290
Qy 345 SVPHPKYIDISDMTISIRLVKMOGHEGIDKCAPYAPCSGTHQRAICVYAGADLMMMLQ 404
Db 291 S-----LNFEHQIMFESG-----CHNMKGSCVFGIGDVSNLQ 327
Qy 405 HHLANKFDPKVDNALOC 424
Db 328 KALVAHKFYLKSEPPAYFCL 347

RESULT 6
T20207

hypothetical protein C54C8.11 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C/Accession: T20207; T27478

R/Dobson, R.
submitted to the EMBL Data Library, November 1996

A/Reference number: Z19237

A/Accession: T20207

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-425 <M12>

A/Cross-references: EMBL:283102; PTDN:CA05469.1; GSPDB:GN00019; CESP:C54C8.11

A/Experimental source: clone C54C8

R/Mclay, K.
submitted to the EMBL Data Library, October 1998

A/Reference number: Z20372

A/Accession: T27478

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-425 <M12>

A/Cross-references: EMBL:AL032664; PTDN:CA21769.1; GSPDB:GN00019; CESP:C54C8.11

A/Experimental source: clone Y91F4A

C/Genetics:

A/Map position: 1

A/Introns: 29/2; 117/3; 161/3; 195/3; 237/3; 266/3; 295/1; 313/3; 355/1; 374/2

Query Match 14.9%; Score 352; DB 2; Length 425;
Best Local Similarity 28.2%; Pred. No. 2.3e-19;
Matches 111; Conservative 58; Mismatches 167; Indels 58; Gaps 14;

Qy 68 INCSGVTRGDOEAVLQAILNMLEYKKRREPTDTHYLS--LTRDCEHFAERKFIQPLPS 125
Db 41 LECOKMFDGPPVSLARKALFKFD--DREIIAQIILKLSGOENAOCAEFQKIFGFQEPPTS 97
Qy 126 KEEVEPPIASVNIHERKIEFERLIRAVYAPQNYCVHDEKSEPFKEAVKAIISCFPN 185
Db 98 QEELEFFPLATMLVHGDFVQLSLISAIYQPONOFCFLAVDGNSSVEFLGLVRLMSRYGN 157
Qy 186 V--FIASKLVAVYASMSRVOADLQSSVPMKYFLNTCGTDPPIKSNMVOAL 243
Db 158 IQYFTLDE--IRMGVEIILTSVFCVDYLAKLPDSMKYFYGLSGVDAPLKSNIEMIRIL 214
Qy 244 KMLNGRSMSESEVPPKHKETRMKYHEVVRDTHLTKKKDPPRYMLTJMTGNATYVASR 303
Db 215 KALNG--SFNAELIP-----FEFYR-----LNKR--PMSSPLPYKTSLSATFSR 256
Qy 304 DFVOHVLKPNKRSQOLIEWKDYSPDEHLMATLO-----RARMPGSVPN----HPK 351
Db 257 KSNAPVAVSEKVLQIDPLRGTTCADSLMATIAGNPKVYKFSLELPMGCGDAMAMHKN 316
Qy 352 YDISDMTISIRLVKMOGHEGIDK-----APYAPCSGTHQRAICVYAGADL 399
Db 317 Y-----RRTRGKLGKY--GENCKIDNGYYSRYQYVNNAP--VKCGYYRRLSCVGVVDLP 370
Qy 400 WMLONHHLANKFDPKVDNALOCLEEYLKAI 433
Db 371 NLINRHELVAKHLYFSYQPAFMCLVENSROKSM 404

RESULT 7
T24930

hypothetical protein T15D6.3 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T24930

R/Dobson, R.
submitted to the EMBL Data Library, November 1996

A/Reference number: Z19956

A/Accession: T24930

Db 271 KLFNDKNSDYELISGVSVHSSLSRAAVKVLVEVLSIFIDOMNTDYGADQFIIST 330
Oy 336 LORA--RMGPSVPHRPYDIDSMTIARLVKMOGHEDDIDGAPYACSGIHQRAICVY 393
Db 331 FQMSDLMKPMGHFTNECHNDIAITVISRLLAWPIY--LDAHAICA--TGVRHDCIF 386
Oy 394 GAGDLMNMLQNHHLANKFD-----PKYDNNALQCLEEYLRYKAIY 435
Db 387 GIEDER-----AISKEPILNMLPARDYSIIECTAELLHNRTYLG 425

RESULT 10
T21408
hypothetical protein F26D2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T21408
R:McMurray, A.
submitted to the EMBL Data Library, November 1996
A:Accession number: Z19418
A:Reference: T21408
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-467 <WILL>
A:Cross-references: EMBL:Z81513; PIDN:CA04185.2; GSPDB:GN00023; CESP:F26D2.3
A:Experimental source: clone F26D2
C:Genetics:
A:Gene: CESP:F26D2.3
A:Map position: 5
A:Introns: 39/1; 90/3; 194/1; 281/1; 320/2; 356/3; 402/1; 438/3
C:Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 11.3%; Score 267; DB 2; Length 467;
Best Local Similarity 24.8%; Pred. No. 9.1e-13;
Matches 102; Conservative 72; Mismatches 176; Indels 62; Gaps 17;

Oy 42 LESRSQSOYCRNIIYNFLKLPARKSINGVTRGDOEAVLQAILNLEVKKKRPETDT 101
Db 55 IEALNMDYCKDRRF-YCKRPETHVDCRRVLKGD-KAYLOSLLG-----INRIPLIEN 107
Oy 102 HYLSTRDCEHFKARKEFIQPLSKEEVEFPYASVMVIEKIEENFERLLRAYAPQNTYC 161
Db 108 PFLNLTCSAIKRRIIPKSSQFKLMLN---GTAFAFIYADVEFLEKQVASYHQNFPC 164
Oy 162 VAVDEKSPFEKEAVKAIISCPPNVFIASKLVRYVYASRSVQADLNCMEDLLQSSVPMK 221
Db 165 FAVDANSSAEFQRMKALERCLPNVFVLPVTEYSYDSKGNIMLAHYNCMK-RLSEASRGWG 223
Oy 222 YFLNTCGTDFPIKSNAEVQALMKLNGRNS--MESEVP-PKHKEPRMKYHFEVVDLHL 278
Db 224 YIMLLQNHVDITKSYELDRIFDLAGANDVPMKSEIPERRKKHSHKMDL-----KSLRL 277
Oy 279 TNKKKDPPEYNLTMTFTGNAYIVASRPFVQHVILKNPSQOLIEWVD-----TY 326
Db 278 FPNESRNSDYELISG-----FVQASL-----SRAVEMLVNEDVSTILLQOPREY 324
Oy 327 SPDEHLMATLQAR--WMPGSVPHRPKYDISMTIARLVKMOGHEDIDGAPYAPCSG 384
Db 335 GADQFISTFQSNQDFKMPGHFTNECFH--NDTSKIFRPSRWV--PINLERSNCA--SG 378
Oy 385 IHQRAICVYGADLMNMLQNHHLANKFDP-KVDNNAQCLEEYLRYKAIY 435
Db 379 TVRHDCIFGIEDF-----LAVSKYPTLTDFSIIECTAELLHNRTYLG 422

RESULT 11
F89364
protein F26D2.3 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: F89364

R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating b1o
A:Reference number: A75000; MUID:99065613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_...
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: F89364
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-483 <STO>
A:Cross-references: GB:chr_V; PIDN:CA04185.1; PID:93924744; GSPDB:GN00023; CESP:F26D2.3
C:Genetics:
A:Gene: F26D2.3
A:Map position: 5
C:Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 11.3%; Score 267; DB 2; Length 483;
Best Local Similarity 24.8%; Pred. No. 9.6e-13;
Matches 102; Conservative 72; Mismatches 176; Indels 62; Gaps 17;

Oy 42 LESRSQSOYCRNIIYNFLKLPARKSINGVTRGDOEAVLQAILNLEVKKKRPETDT 101
Db 71 IEALNMDYCKDRRF-YCKRPETHVDCRRVLKGD-KAYLOSLLG-----INRIPLIEN 123
Oy 102 HYLSTRDCEHFKARKEFIQPLSKEEVEFPYASVMVIEKIEENFERLLRAYAPQNTYC 161
Db 124 PFLNLTCSAIKRRIIPKSSQFKLMLN---GTAFAFIYADVEFLEKQVASYHQNFPC 180
Oy 162 VAVDEKSPFEKEAVKAIISCPPNVFIASKLVRYVYASRSVQADLNCMEDLLQSSVPMK 221
Db 181 FAVDANSSAEFQRMKALERCLPNVFVLPVTEYSYDSKGNIMLAHYNCMK-RLSEASRGWG 239
Oy 222 YFLNTCGTDFPIKSNAEVQALMKLNGRNS--MESEVP-PKHKEPRMKYHFEVVDLHL 278
Db 240 YIMLLQNHVDITKSYELDRIFDLAGANDVPMKSEIPERRKKHSHKMDL-----KSLRL 293
Oy 279 TNKKKDPPEYNLTMTFTGNAYIVASRPFVQHVILKNPSQOLIEWVD-----TY 326
Db 294 FPNESRNSDYELISG-----FVQASL-----SRAVEMLVNEDVSTILLQOPREY 340
Oy 327 SPDEHLMATLQAR--WMPGSVPHRPKYDISMTIARLVKMOGHEDIDGAPYAPCSG 384
Db 341 GADQFISTFQSNQDFKMPGHFTNECFH--NDTSKIFRPSRWV--PINLERSNCA--SG 394
Oy 385 IHQRAICVYGADLMNMLQNHHLANKFDP-KVDNNAQCLEEYLRYKAIY 435
Db 395 TVRHDCIFGIEDF-----LAVSKYPTLTDFSIIECTAELLHNRTYLG 438

RESULT 12
D87603
glycosyl transferase, probable CC2860 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: D87603
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Neilson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Ko n, J.; Emdolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: D87603
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-322 <STO>
A:Cross-references: GB:A8005673; NID:913424474; PIDN:AAK24824.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2860

Query Match 11.0%; Score 259.5; DB 2; Length 322;
Best Local Similarity 26.1%; Pred. No. 2.1e-12;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 09:45:11 : Search time 19.6 Seconds

(without alignments)
865.263 Million cell updates/sec

Title: US-09-874-390-2

Perfect score: 2360

Sequence: 1 MYQWKRLLCQLHYLWALGICYM.....NALQCLEYLRYKAYGTSL 438

Scoring table: BLOSUM62

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1142.5	48.4	428 1	GGNT_HUMAN
2	1111.5	47.1	428 1	GGNT_MOUSE
3	1106	46.9	427 1	GGNT_BOVIN
4	801.5	34.0	400 1	BGIB_MOUSE
5	784	33.2	400 1	BGIB_HUMAN
6	104.5	4.4	300 1	EXRN_BPT7
7	104.5	4.4	560 1	INRL_SHEEP
8	101.5	4.3	302 1	EXRN_BPT3
9	100.5	4.3	2555 1	FAFY_HUMAN
10	100	4.2	832 1	TRNL_CANAL
11	97	4.1	987 1	TRPA_BACTU
12	96.5	4.0	779 1	MSS4_YEAST
13	94.5	4.0	731 1	UR87_HSVSA
14	94	4.0	411 1	THBG_BOVIN
15	93.5	4.0	879 1	RA50_PYROO
16	92.5	3.9	619 1	LEU1_YEAST
17	92	3.9	992 1	PHS2_DICDI
18	91.5	3.9	738 1	CHSY_USTMA
19	91.5	3.9	960 1	KRT_CHICK
20	91	3.9	2663 1	CENE_HUMAN
21	91	3.9	3135 1	S230_PLAFO
22	90.5	3.8	876 1	RPOB_NPVAC
23	90.5	3.8	1013 1	CHSA_EMENT
24	90.5	3.8	1018 1	YKC2_CAEEL
25	90.5	3.8	1230 1	UBPF_YEAST
26	90.5	3.8	2559 1	TRFX_MOUSE
27	90	3.8	669 1	MYBE_AVILE
28	90	3.8	1514 1	YD34_YEAST
29	89.5	3.8	354 1	YSCU_YERPE
30	89.5	3.8	1597 1	GMFL_STROO
31	89.5	3.8	4969 1	RRR2_RABIT
32	89	3.8	1717 1	RRP1_HUMAN
33	89	3.8	3744 1	YHP9_YEAST

34	89	3.8	6669 1	NEBU_HUMAN	P20929 homo sapien
35	88.5	3.8	515 1	CP51_PENIT	Q12664 penicillium
36	88.5	3.8	3068 1	POLG_PEMVC	001500 p genome po
37	88.5	3.8	5430 1	ACF7_HUMAN	09upn3 homo sapien
38	88	3.7	578 1	YWRJ_CAEEL	Q20471 caenorhabdi
39	88	3.7	601 1	YBM4_SCHPO	Q10338 schizosacch
40	88	3.7	996 1	PIP2_YEAST	P52960 saccharomyc
41	87.5	3.7	522 1	N162_HUMAN	P37198 homo sapien
42	87.5	3.7	540 1	AXR1_ARATH	P42744 arabidopsis
43	87	3.7	1011 1	VPS3_YEAST	P23643 saccharomyc
44	86.5	3.7	359 1	DNLI_BPT7	P00969 bacterioph
45	86.5	3.7	824 1	DPOL_METVO	P52025 methanococc

ALIGNMENTS

```

RESULT 1
ID      G6NT_HUMAN      STANDARD:      PRT:      428 AA.
AC      002742:
DT      01-JUN-1994 (Rel. 29, Last sequence update)
DT      01-JUN-1994 (Rel. 29, Last sequence update)
DT      01-MAR-2002 (Rel. 41, Last annotation update)
DE      Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-
DE      acetylglucosaminyltransferase (EC 2.4.1.102) (Core 2 branching enzyme)
DE      (Core2-GlcNAc-transferase) (C2GNT) (Core 2 GNT).
CN      GCM1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93028457; PubMed=1329093;
RA      Bierhuizen M.F.A., Fukuda M.;
RT      "Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta
RT      1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6GlcNAc transferase by gene
RT      transfer into CHO cells expressing polyoma large tumor antigen.";
RL      Proc. Natl. Acad. Sci. U.S.A. 89:9326-9330(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=96078409; PubMed=7579796;
RA      Bierhuizen M.F.A., Maemura K., Kudo S., Fukuda M.;
RT      "Genomic organization of core 2 and I branching beta-1,6-N-
RT      acetylglucosaminyltransferases. Implication for evolution of the
RT      beta-1,6-N-acetylglucosaminyltransferase gene family.";
RL      Glycobiology 5:417-425(1995).
CC      -1- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.
CC      -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + beta-D-
CC      galactosyl-1,3-(N-acetyl-beta-D-glucosaminyl-1,6)-N-acetyl-D-
CC      galactosyl-1,3-(N-acetyl-beta-D-glucosaminyl-1,6)-N-acetyl-D-
CC      galactosaminyl-R.
CC      -1- PATHWAY: GLYCOSYLATION.
CC      -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
CC      -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ACTIVATED T-LYMPHOCYTES
CC      AND MYELOID CELLS.
CC      -1- SIMILARITY: TO I-BRANCHING ENZYME (GNT).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M97347; AAA5919.1; -.
CC      EMBL: L41415; AAA96661.1; -.
CC      PIR: A46293; A46293.
CC      MIM: 600391; -.
CC      InterPro: IPR003406; Branch.

```

[illegible]

RN	J. Biol. Chem.	272:27246-27252(1997).
RL	[3]	
RP	SEQUENCE OF 62-72; 109-116; 225-232 AND 258-265, AND FUNCTION.	
RC	TISSUE=Kidney;	
RX	MEDLINE=95074155; PubMed=7983056;	
RA	Sekine M., Hashimoto Y., Suzuki M., Inagaki F., Takio K., Suzuki A.;	
RT	"Purification and characterization of UDP-glucNAc:UV3 beta Gal-GlcAcer beta-1,6-glcNAc transferase from mouse kidney.";	
RT	J. Biol. Chem. 269:31143-31148(1994).	
CC	-1- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.	
CC	-1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + beta-D-	
CC	galactosyl-1,3-N-acetyl-D-galactosaminyl-R = UDP + beta-D-	
CC	galactosyl-1,3-(N-acetyl-beta-D-glucosaminyl-1,6)-N-acetyl-D-	
CC	galactosaminyl-R.	
CC	-1- PATHWAY: GLYCOSYLATION.	
CC	-1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.	
CC	-1- TISSUE SPECIFICITY: Expressed in kidney, liver, stomach, spleen, lung and brain.	
CC	-1- SIMILARITY: TO I-BRANCHING ENZYME (IGNT).	
CC	-----	
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration - between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on ways use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license@lsb.scrib.ch).	
CC	or send an email to license@lsb.scrib.ch).	
CC	-----	
DR	EMBL; U19265; AAA60948.1; -	
DR	EMBL; D87333; BAA22999.1; -	
DR	EMBL; D87332; BAA22998.1; -	
DR	GSD; MG1:95676; GenCl.	
DR	InterPro; IPRO03406; Branch.	
DR	Pfam; PF02485; Branch; 1.	
KW	Transferase; Glycosyltransferase; Transmembrane; signal-anchor;	
KW	Golgi stack; Glycoprotein.	
FT	DOMAIN	1 9 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	10 32 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT		(POTENTIAL).
FT	DOMAIN	33 428 LUMENAL, CATALYTIC (POTENTIAL).
FT	CARBOND	58 58 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOND	95 95 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	191 191 T->S (IN REF. 2).
FT	CONFLICT	238 238 G->A (IN REF. 2).
FT	CONFLICT	404 404 M->I (IN REF. 2).
SO	SEQUENCE	428 AA; 49849 MW; F3FFD251B27AAE1 CRC64;
<hr/>		
Query Match	47.1%; Score 1111.5; DB 1; Length 428;	
Best Local Similarity	54.8%; Pred. NO.1.5e-77;	
Matches 204; Conservative	65; Mismatches 102; Indels 1; Gaps	1;
OY	67 SINCSCVGRGOAEVALQALLNLNLEVK-KKREPTGTHYLSLRDCEHFKAEKKFIOTPLS	125
Ddb	::::: : :: :	115
OY	126 KEVEEPFIAYSMWJHEKIETNEFERLLRAVAYAPONIVCVHNDEKSPTPEAKVAKAISCPEN	185
Ddb	:::	175
OY	116 KEEVEFPFIAYSIVNHKITEMDLRLRALATYPMPNCIHVDKRAEESFLAAVOGIASCFDN	175
Ddb	: :	245
OY	186 VFIAASKLYRVVYASRSVQADLNCEDELQSSVPVKYFLNTCGTFPPIKSAENMVOQLKM	245
Ddb	: :	235
OY	176 VFAVASLESVYASVTBYKADLNCCKDKDIRMNANKRYLINLCGMDFPIKTNDIEIVRKLIK	235
Ddb	: :	305
OY	246 LNGRSNMESSEVPRIKKEERHWKYHEFYVADVLTJLTKKKDKDPREYNLMFGNGNYIASRDE	305
Ddb	: :	295
OY	306 VOHVIAKNKSOOLIMWVADTYSPDEHLATLQARAAMMGGSVNPBKHYDISDTSTARLK	365
Ddb	: :	355
OY	366 WOGHEGDIDKGAPYPCSGIHORALICYVGAGDJNMMLQNHLANKPDPKVDNALQCLE	425

Db 356 WQYFEGDVSNGAPRPPCSGVHRSVCVFAGDLSWMLRQHHLFANKFDMDDVPFAIQCD 415
 QY 426 EYLRYKATGTE 437
 Db 416 EHLRRKALENLE 427

RESULT 3

GENT_BOVIN

ID G6NT_BOVIN STANDARD: PRT: 427 AA.

AC 092180.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase (EC 2.4.1.102) (core 2 branching enzyme) (CORE2-GLCNAC-transferase) (C2GNT) (core 2 GNT).
 GN GCNT1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.
 NCBI_TaxID=9913;

CC -1- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + beta-D-galactosyl-1,3-N-acetyl-D-galactosaminyl-R = UDP + beta-D-galactosaminyl-1,3-(N-acetyl-beta-D-glucosaminyl-1,6)-N-acetyl-D-galactosaminyl-R.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
 CC -1- SIMILARITY: TO I-BRANCHING ENZYME (IGNT).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----

DR EMBL: U41320: AAA83244.1: -
 DR InterPro: IPR003406: Branch.
 DR Pfam: PF02485: Branch: 1.
 KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor; Golgi stack; Glycoprotein.
 FT DOMAIN 1 9
 FT TRANSMEM 10 32
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 33 427
 FT CARBOHYD 52 52
 FT CARBOHYD 58 58
 FT CARBOHYD 95 95
 FT SEQUENCE 427 AA: 49733 MW: 4189CBFD948D4196 CRC64:

Query Match Best Local Similarity 46.9%; Score 1106; DB 1: Length 427; Matches 204; Conservative 64; Mismatches 99; Indels 2; Gaps 2;

QY 67 SINCSGVTRGDQDAVLAQALNNLEVK-KKREPTDTHTLSLRDCEHFAKRRKFIQPLS 125
 Db 56 NINCKRIILQGDVDEIQKVLKSLTGVKKRARTNVDINMTGCASFIKKRYITEPLS 115
 QY 126 KEVEEPPIAIVNHIKTEIFRLRAYAAPONICVAHDESPETFEKAVAIISCFPN 185
 Db 116 KEEAGPIAIVNHIKTEIFRLRAYAAPONICVAHDESPETFEKAVAIISCFPN 175

QY 186 VEIASKLVRYVYASNSRVOADLNCMEDLQSSVPMKFLNTGCTDPEPKSNAEMVOALKM 245
 Db 176 VFAVAGLSLEVVYASNSRVOADLNCMODLVQNMAGMKYLLNLGCMDEPITNLEIYRKLL 235
 QY 246 LMGNSMSESEVPPKHETRWKTHFEVYRDTLTLTNKKDPPRYNLTFPGNAIYASRDF 305
 Db 236 LMGENNLETERKMPSHKKEKRWKHYEVVNGKLTNMGTDKIHPLPLEPLFGSAHFVVSREY 295
 QY 306 VOHVLKNPKSQQLIEKVTYSPDEHLMATLGRARHMPGCVPNHHPYDIDSMTSIARLVK 365
 Db 296 VETVLQNONIQFMEWAKDTYSPDEHLMATLGRARHMPGCVPNHHPYDIDSMTSIARLVK 355
 QY 366 WQHGEGDIDKAPYAPCSGIRHORAIQVCGADLNNMLQNHHLANKFDPKVDNALQCLCE 425
 Db 356 WQYFEGDVSNGAPRPPCS-VHRSVCVFAGDLSWMLRQHHLFANKFDMDDVPFAIQCD 414
 QY 426 EYLRYKAT 433
 Db 415 EHLRRKAL 422

RESULT 4

BGTB_MOUSE

ID BGTB_MOUSE STANDARD: PRT: 400 AA.

AC P97402.
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE N-acetylglucosaminide beta-1,6-N-acetylglucosaminyltransferase (EC 2.4.1.150) (N-acetylglucosaminyltransferase) (I-branching enzyme) (IGNT) (large I antigen-forming beta-1,6-N-acetylglucosaminyltransferase).
 GN GCNT2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;

CC -1- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR BRANCHED POLY-N-ACETYLGLUCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + beta-D-galactosyl-1,4-N-acetyl-D-glucosaminyl-R = UDP + N-acetyl-D-glucosaminyl-1,6-beta-D-galactosyl-1,4-N-acetyl-D-glucosaminyl-R.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
 CC -1- SIMILARITY: TO CORE 2 BRANCHING ENZYME (C2GNT).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----

DR EMBL: U68182: AAB39621.1: -
 DR MGD: MGI:1100870: Gcnt2.
 DR InterPro: IPR003406: Branch.
 DR Pfam: PF02485: Branch: 1.
 KW Transferase; Glycosyltransferase; Transmembrane; Signal-anchor; Golgi stack; Glycoprotein.
 FT DOMAIN 1 6
 FT TRANSMEM 7 25
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT CYTOPLASMIC (POTENTIAL).
 FT LUMENAL, CATALYTIC (POTENTIAL).

FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 400 AA: 45507 MW: 193FF3882E3B9691 CRC64;

Query Match 34.0%; Score 801.5; DB 1; Length 400;
 Best Local Similarity 43.2%; Pred. No. 6e-54;
 Matches 162; Conservative 59; Mismatches 129; Indels 25; Gaps 5;

QY 76 GDOEAVLQALNNLEVKKKRPPTDHYLSLTRD-----CEHKARKKFIQFPLSK 126
 DB 29 GDSYQKLNISDSVRLSOVCSSFDGKSRLPLMKMLIHEKPSCTEVTQSHYITAPLSQ 88
 QY 127 EEEVFPYASVNIHEKIEENFRLRAYAPONTICVHDEKSPETFEAVAKISCPNV 186
 DB 89 EEGGFPLAYVNIHNFDTARLFRAIFPMONITCYHDEKATREFGAVDQLVSCFPNV 148
 QY 187 FIASKLVVVYASWSRVADLNCMEDLLQSSVPKRYFLNTGCDPPIKSNAEMVQALKML 246
 DB 149 FIASKMPEVYVGGISRLQADLNCIKIDLTSEVPKRYAINTCGQDFPIKTKKEIYQYLKGL 208
 QY 247 NGRSMSEVPPKKETRMKY-HFEVVRDTHL---TNKKKDPPIVULMTGNAIYAS 302
 DB 209 KKKMLPFGVLPRAHAIQRTYRHEHLSKELSYIRITALKPPEPHMLTTFGSAVYALS 268
 QY 303 RDEYVHVALPKPSQQLIEYKDYSPDEHMLTQARARMPGSPVNPXYDISMTSIAR 362
 DB 269 REFANFPLRPRAVDLHWSKDFSPDEHFWTLNRIIPGSPPN-----ASWTGNLR 322
 QY 363 LVKQGHGCDIDGAPAPCSGTHQRAICVYAGADLNMMLQNHHLANKCDPKVDDNALQ 422
 DB 323 AVKM-----MDMKAKHGCGQGHVHGICITGNMDLQMLINSOSLFANKKELNTYPLTVE 376
 QY 423 CLEERYLKRYAIGTE 437
 DB 377 CLERLRLERLINOSE 391

RESULT 5
 BGIB_HUMAN
 ID BGIB_HUMAN STANDARD; PRT; 400 AA.
 AC 006430;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE N-acetylglucosaminide beta-1,6-N-acetylglucosaminyltransferase
 DE (EC 2.4.1.150) (N-acetylglucosaminyltransferase) (I-branching enzyme)
 DE (IGMT).
 GN GCNT2 OR IT.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93194065; PubMed=8449405;
 RA Bierhuizen M.F.A., Mattei M.-G., Fukuda M.;
 RT "Expression of the developmental I antigen by a cloned human cDNA
 RT encoding a member of a beta-1,6-N-acetylglucosaminyltransferase gene
 RT family.";
 RL Genes Dev. 7:468-478(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=96078409; PubMed=7579796;
 RA Bierhuizen M.F.A., Maemura K., Kudo S., Fukuda M.;
 RT "Genomic organization of core 2 and I branching beta-1,6-N-
 RT acetylglucosaminyltransferases. Implication for evolution of the
 RT beta-1,6-N-acetylglucosaminyltransferase gene family.";
 RL Glycobiology 5:417-425(1995).
 RN [3]

RP TISSUE SPECIFICITY.
 RX MEDLINE=98070745; PubMed=9405606;
 RA Sasaki K., Kureta-Miura K., Ujita M., Angata K., Nakagawa S.,
 RA Sekine S., Nishi T., Fukuda M.;
 RT "Expression cloning of cDNA encoding a human beta-1,3-N-
 RT acetylglucosaminyltransferase that is essential for poly-N-
 RT acetylglucosamine synthesis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:14294-14299(1997).
 CC -I- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR BRANCHED
 CC POLY-N-ACETYLGLUCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I
 CC ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED
 CC WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.
 CC -I- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + beta-D-
 CC galactosyl-1,4-N-acetyl-D-glucosaminyl-R = UDP + N-acetyl-beta-D-
 CC glucosaminyl-1,6-beta-D-galactosyl-1,4-N-acetyl-D-glucosaminyl-R.
 CC -I- PATHWAY: GLYCOSYLATION.
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Golgi.
 CC -I- TISSUE SPECIFICITY: IN THE ADULT, HIGHLY EXPRESSED IN PROSTATE AND
 CC TO A LESSER EXTENT IN SMALL INTESTINE AND COLON. BARELY DETECTED
 CC IN HEART, BRAIN, KIDNEY AND PANCREAS. NO EXPRESSION IN PLACENTA,
 CC LUNG, LIVER, SKELETAL MUSCLE, SPLEEN, THYMUS, TESTIS, OVARY AND
 CC PERIPHERAL BLOOD LEUKOCYTES. IN FETUS, HIGHLY EXPRESSED IN BRAIN
 CC AND TO A LESSER EXTENT IN LUNG AND KIDNEY. BARELY DETECTED IN
 CC LIVER.
 CC -I- DEVELOPMENTAL STAGE: ITS EXPRESSION INCREASES DRAMATICALLY DURING
 CC -I- SIMILARITY: TO CORE 2 BRANCHING ENZYME (C2GNT).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: 119659; AAA81777.1; -
 DR EMBL: 219550; CAAT9610.1; -
 DR EMBL: 141607; AAA99832.1; -
 DR EMBL: 141605; AAA99832.1; JOINED.
 DR EMBL: 141606; AAA99832.1; JOINED.
 DR PIR: A46297; A46297.
 DR MIM: 600429; -
 DR MIM: 110800; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 KW Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
 KW Golgi stack; Glycoprotein.
 FT DOMAIN 1 6
 FT TRANSMEM 7 25
 FT
 FT DOMAIN 26 400
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 400 AA: 45854 MW: 1469628690A1F43B CRC64;

Query Match 33.2%; Score 784; DB 1; Length 400;
 Best Local Similarity 42.3%; Pred. No. 1.3e-52;
 Matches 163; Conservative 62; Mismatches 126; Indels 34; Gaps 9;

QY 56 LYNFLKLPARKSINGSGVTRGDEAVLQALNNLEVKKKRPPTDHYLSLTRDCEHFA 115
 DB 24 VFNNGDPSFQRLNISDPLRLTQ--VCTSPING---KTRFLMKMKMLIHEKSSCKEYLT 77
 QY 116 ERKFIQPLSKSEVEPPIASVNIHEKIEENFRLRAYAPONTICVHDEKSPETFEKA 175
 DB 78 QSHYITAPLSKEADPLAYVNIHNFDTARLFRAIFPMONITCYHDEKATTEKKA 137
 QY 176 VKAIIISCFNVFIASKLVVRVYASWSRVADLNCMEDLLQSSVPKRYFLNTGCDPPIKS 235
 I: ::::: 1:1111: 111 11:111111: 11 111:1111 11111:

Db 138 VEQLLSCFPNMFASKMEPVYVGGISRLQADINCRJDSAFEVSKYVINTCGDPPLEKT 197
 QY 236 NAEWYQALKMLNGRNSMESEVPPKHETRWKY-HFE-----VVRDTLHLTNKKDDP 286
 Db 198 NKEIYQYLKGFGRKNTITGVLPRAHAIGRTKYVHOEHLGKELSYIR-----TTLKKRPP 252
 QY 287 PYNLTFMGNAVITYASRDPVOHVHLKNPKSQQLIEWKDTYSPDEHLMATLQARMMPGSV 346
 Db 253 PHNLTYFGSAVVALSRFANFVLHDPRAVDLQWSKDTFSPDEHFWTLNRPVPGSGM 312
 QY 347 PNHPRYDLSDMTSTARLYKMQGHEGDDIDGAPYACSGIHQALTCVYGAGDLMMLQNH 406
 Db 313 PN-----ASWTGNLRATIKWSDME--DR--HGCGHGHVYHICITYGNGDLKMLNPS 360
 QY 407 LLANKFDPKVDNALQCLEEYLRK 431
 Db 361 LEANKFELNTPYIVECLE--LRHR 383
 RESULT 6
 EXRN_BPT7 STANDARD: PRT: 300 AA.
 AC P00638:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Exodeoxyribonuclease (EC 3.1.11.3) (Exonuclease).
 GN 6.
 OS Bacteriophage T7.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
 CC T7-like phages.
 OX NCBI_TaxID=10760;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83241725; PubMed=6864790;
 RA Dunn J.J., Studier F.W.;
 RT "Complete nucleotide sequence of bacteriophage T7 DNA and the
 RL locations of T7 genetic elements.";
 RL J. Mol. Biol. 166:477-535(1983).
 RN [2]
 RP REVISION.
 RX MEDLINE=90133923; PubMed=2614843;
 RA Dunn J.J.;
 RL Unpublished results, cited by:
 RL Beck P.J., Gonzalez S., Ward C.L., Molineux I.J.;
 RL J. Mol. Biol. 210:687-701(1989).
 CC -1- FUNCTION: This enzyme is essential for phage DNA replication; it
 CC is believed to function in the removal of DNA-linked RNA primers.
 CC It is also necessary for host DNA degradation and phage genetic
 CC recombination.
 CC -1- CATALYTIC ACTIVITY: Degradation of double-stranded DNA. It acts
 CC progressively in a 5'- to 3'-direction, releasing 5'-
 CC phosphomononucleotides.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: V01146; CAA24418.1; .
 DR PIR: A00780; NDBPT7.
 DR PIR: S42316; S42316.
 DR InterPro: IPR000513; Exo_N.I.
 KW Hydrolyase; Nuclease; Exonuclease.
 SQ SEQUENCE 300 AA; 34502 MW; 750BE36D527E31C4 CRC64;

Query Match 4.48; Score 104.5; DB 1; Length 300;
 Best Local Similarity 19.78; Pred. No. 0.58;
 Matches 62; Conservative 46; Mismatches 107; Indels 99; Gaps 14;

QY 81 VLQATLNNLEVKK-----REPTDTHYLSLTRDCEHFAERKEFIQF 123
 Db 60 ILEDISIKSEYERKKAMAGAPIVLAFDVSVMKKELVDPNY-----KANRR----- 104
 QY 124 LSKEEVEEPPIASVMVIEKIEFNERLLRAVYAPQNIYCVHDEKSPETFEKAVKAIISCF 183
 Db 105 ----AVKKRPVGY-----FE-FLDALFEREEFYCI---RPMLEGDDVQVGIASN 145
 QY 184 PVVFIAKSLVRYVYASWSVQADLNCMEDLLQSSVPWKYFLNTGCTDPPIKSNAMVQAL 243
 Db 146 PSAFGARRAVII-----SCDKRFTIPNCF---L 172
 QY 244 KMLNGRNSMESEVPPKHETRWKYHFEVVRDTLHLTNKKDDPPYNLFW--FTGNAYIVA 301
 Db 173 WCTTGNILQTE-----ESADWMHILFQITKG--DITDQSGIAGGDTADPEDLNNPFI 225
 QY 302 SRDPVOHVHLKNPKSQQLIEWKDTYSPDEHLMATLQARMMPGSVNHPKDYISDMTSTA 361
 Db 226 PRTSVLKSGKN-KGOEVIKWKVRDEPHETLMDCKSI---GAKAGMTEEDIIKQGWMA 280
 QY 362 RLVMQGHEGDDK 375
 Db 281 RLRFNEY-NFLDK 293
 RESULT 7
 INRL_SHEEP STANDARD: PRT: 560 AA.
 AC Q28589; Q95206;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Interferon-alpha/beta receptor alpha chain precursor (IFN-alpha-REC)
 DE (interferon alpha/beta receptor-1).
 GN IFNAR1 OR IFNAR.
 OS Ovis aries (Sheep).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Caprinae; Ovis.
 OX NCBI_TaxID=9940;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Endometrium;
 RL MEDLINE=97135690; PubMed=8981227;
 RA Kaluz S., Fisher P.A., Kaluzova M., Sheidrick E.L., Flint A.P.F.;
 RT "Structure of an ovine interferon receptor and its expression in
 RL endometrium.";
 RL J. Mol. Endocrinol. 17:207-215(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Endometrium;
 RL MEDLINE=98006426; PubMed=9348203;
 RA Han C.-S., Mathalagan N., Klemann S.W., Roberts R.M.;
 RT "Molecular cloning of ovine and bovine type I interferon receptor
 RT subunits from uteri, and endometrial expression of messenger
 RT ribonucleic acid for ovine receptors during the estrous cycle and
 RT pregnancy.";
 RL Endocrinology 138:4757-4767(1997).
 RL -1- FUNCTION: RECEPTOR FOR INTERFERONS ALPHA AND BETA. BINDING TO TYPE
 CC I IFNS TRIGGERS TYROSINE PHOSPHORYLATION OF A NUMBER OF PROTEINS
 CC INCLUDING JAKS, TYK2, STAT PROTEINS AND IFN-R ALPHA-AND BETA-
 CC SUBUNTS THEMSELVES.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED EXCEPT
 CC CONCEPTUS AT DAY 15 OF PREGNANCY.
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE CLASS II CYTOKINE FAMILY OF RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI Taxid=9606;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Testis:
 RX MEDLINE=97081753; PubMed=8922996;
 RA Jones M.H., Furlong R.A., Burkin H., Chalmers I.J., Brown G.M.,
 RA Khwaja O., Affara N.A.;
 RT "The Drosophila developmental gene fat facets has a human homologue
 RT in Xp11.4 which escapes X-inactivation and has related sequences on
 RT Yq11.2.";
 RL Hum. Mol. Genet. 5:1695-1701(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Brown G.M., Sargent C.A., Erickson R.P., Mazayrat S.,
 RA Mitchell M.J., Jones M.H., Cooke H., Affara N.A.;
 RL Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98022381; PubMed=9381176;
 RA Iain B.T., Page D.C.;
 RT "Functional coherence of the human Y chromosome.";
 RL Science 278:675-680(1997).
 RN [4]
 RP SEQUENCE FROM N.A., AND GENE MAPPING.
 RC TISSUE-Testis, Fetal brain, and Retina;
 RX MEDLINE=98046012; PubMed=9384609;
 RA Brown G.M., Furlong R.A., Sargent C.A., Erickson R.P., Longepied G.,
 RA Mitchell M., Jones M.H., Hargreave T.B., Cooke H.J., Affara N.A.;
 RT "Characterisation of the coding sequence and fine mapping of the human
 RT Sxrb gene and comparative expression analysis and mapping to the
 RT Sxrb interval of the mouse Y chromosome of the Dfry gene.";
 RL Hum. Mol. Genet. 7:97-107(1998).
 CC -1- FUNCTION: MAY FUNCTION AS AN UBIQUITIN-PROTEIN OR POLYUBQUITIN
 CC HYDROLASE INVOLVED BOTH IN THE PROCESSING OF UBIQUITIN PRECURSORS
 CC AND OF UBIQUITINATED PROTEINS. MAY THEREFORE PLAY AN IMPORTANT
 CC ROLE REGULATORY ROLE AT THE LEVEL OF PROTEIN TURNOVER BY
 CC PREVENTING DEGRADATION OF PROTEINS THROUGH THE REMOVAL OF
 CC CONUGATED UBIQUITIN.
 CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
 CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN EMBRYONIC AND ADULT
 CC TISSUES.
 CC -1- DISEASE: DEFECTS IN USP9Y MAY BE THE CAUSE OF MALE INFERTILITY OF
 CC THE AZOOSPERMIA FACTOR A (AZFA) SPERMATOGENIC PHENOTYPE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS
 CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: Y13618; CAA73940.1; -;
 DR EMBL: Y13619; CAA73941.1; -;
 DR EMBL: AF000986; AAC51833.1; -;
 DR MEROPS: C19.028; -;
 DR MIM: 400005; -;
 DR InterPro: IPR001394; UCH-2.
 DR Pfam: PF00442; UCH-1; 1.
 DR Pfam: PF00443; UCH-2; 1.
 DR PROSITE: PS00972; UCH_2_1; 1.
 DR PROSITE: PS00973; UCH_2_2; 1.
 DR PROSITE: PS50235; UCH_2_3; 1.
 KW Ubiquitin conjugation; Hydrolase; Thiol protease; Multigene family;
 KW Alternative splicing.
 FT ACT_SITE 1568 1568 BY SIMILARITY.
 FT ACT_SITE 1873 1873 BY SIMILARITY.

FT ACT_SITE 1881 1881 BY SIMILARITY.
 FT VARSPLIC 2071 2555 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 206 206 E -> D (IN REF. 3).
 FT CONFLICT 282 282 I -> M (IN REF. 3).
 FT CONFLICT 524 524 D -> Y (IN REF. 3).
 FT CONFLICT 542 542 S -> A (IN REF. 3).
 FT CONFLICT 666 666 L -> V (IN REF. 3).
 FT CONFLICT 883 883 C -> R (IN REF. 3).
 FT CONFLICT 907 907 F -> S (IN REF. 3).
 FT CONFLICT 946 946 N -> D (IN REF. 3).
 FT CONFLICT 1002 1002 K -> E (IN REF. 3).
 FT CONFLICT 1017 1017 K -> R (IN REF. 3).
 FT CONFLICT 1025 1027 FAN -> VAD (IN REF. 3).
 FT CONFLICT 1038 1039 KN -> RD (IN REF. 3).
 FT CONFLICT 1060 1060 T -> A (IN REF. 3).
 FT CONFLICT 1067 1067 N -> K (IN REF. 3).
 FT CONFLICT 1258 1258 F -> V (IN REF. 3).
 FT CONFLICT 1733 1733 F -> Y (IN REF. 3).
 FT CONFLICT 1953 1953 P -> L (IN REF. 3).
 FT CONFLICT 2086 2086 R -> G (IN REF. 3).
 SQ SEQUENCE 2555 AA; 291115 MW; 8DEC03DF0472B995 CRC64;
 Query Match 4.3%; Score 100.5; DB 1; Length 2555;
 Best Local Similarity 19.4%; Pred. No. 19;
 Matches 102; Conservative 72; Mismatches 189; Indels 163; Gaps 24;
 OY 18 CYMLATYALKIS-----FRKCDSHLGLESRESOYCR----- 53
 DB 1297 GFALPTALDASKERAMQTFIIDLILNPKSTVQRLAOEOPFLRCRCMGHRLPEFI 1356
 OY 54 NLVNFELPKAPKRSNGSVTRGDQAVLYALTN-----NLEV-----KKK 94
 DB 1337 TLFFTLIGSTARE-----KQYSGDYFTLLRLNLAYNGNINIPNAEVLVSEIMLKRI 1412
 OY 95 R-----EFTDTHVLSLRDCEHFAKRRKFIQPLSKBEVEPIYASWVHER 142
 DB 1413 RNVKANTGETGVEEPILEGH-LGVTKELLAFQTSKKYHFGCEK-----GGANLIKEL 1464
 OY 143 IENFERLRAYAPONTICVHDEKSPETFKAVYVAIISCFNVTIAS---KLVRYVNAS 199
 DB 1465 IDDF-----IFPASKVLYQL--RSGEL--PAEOAIPVCSSPVITINGFELLVALATGC 1514
 OY 200 WSRVQADLNCMDL-----LOSSVWKYF-----LNTCGDFPFIKSA 237
 DB 1515 VANKIYDCLTEMYTGMTAITTCALTEWEYLPVPGPPRPFGVGLKAGATCTYNS-- 1572
 OY 238 ENVOALKMLNG-RNSM-----ESEVPKHKETRMKRYHEVYARD 274
 DB 1573 -VIQQLYMIPIRSISILAEIGTGSDDLHDMFGDEKQDSNSVNDPRDDVFGYHOFD---D 1628
 OY 275 TLHLTNKKKDPPEYVLTMTFGNATYVASRDEYQVHLKPKESQQLIEWYKDIYSPEHLMA 334
 DB 1629 KPAL-SKTEDEKREYVIGVLR-----HLOVFEGHLASQQLYYVPRGFMKQFRLMG 1677
 OY 335 TLQRRAMPGPSVPNPKVDISMTSIRLVYKMGHGEQIDK--GAPYAP---CSGILHORA 389
 DB 1678 EPNVLRQGHDAI-----EPPNSLVDSLDLAKLKGPAIISKVLGSGFADQKICGCPHRE 1733
 OY 390 ICVYGAGDLNMMALQNHHLANKFDEKVDVDMALOCLEELYRYKAIAG 435
 DB 1734 BCESEFTLLNDVDIRNH-----QNLLDSLEQYIKGDLLEG 1767
 RESULT 10
 TRNL_CANAL
 ID TRNL_CANAL STANDARD: PRT; 832 AA.
 AC P43075:
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE tRNA ligase (EC 6.5.1.3).
 GN TRNL OR RLG1 OR LIG1.

```

OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC5314;
RX MEDLINE=94237478; PubMed=8181747;
RA Baymiller J., Jennings S., Kienzle B., Gorman J.A., Kelly R.,
RT "Isolation and sequence of the t-RNA ligase-encoding gene of Candida
RT albicans."
RT Gene 142:129-134(1994).
CC -1- FUNCTION: ONE OF THE TWO PROTEINS REQUIRED FOR THE SPLICING OF
CC PRECURSOR tRNA MOLECULES CONTAINING INTRONS. THE LIGATION ACTIVITY
CC REQUIRE THREE ENZYMAIC ACTIVITIES: PHOSPHORYLATION OF THE 5'
CC TERMINUS OF THE 3' HALF-tRNA IN THE PRESENCE OF ATP, OPENING OF
CC THE 2',3'-CYCLIC PHOSPHODIESTER BOND OF THE 5' HALF-tRNA LEAVING A
CC 2'-PHOSPHOMONESTER AND LIGATION OF THE TWO tRNA HALVES IN A ATP-
CC DEPENDENT REACTION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + (ribonucleotide)(N) +
CC (ribonucleotide)(M) = AMP + diphosphate + (ribonucleotide)(N+M).
CC -1- DOMAIN: HAS THREE DOMAIN EACH CORRESPONDING TO AN ENZYMAIC
CC ACTIVITY, NAMELY IN N- TO C-TERMINAL ORDER: LIGASE, KINASE AND
CC CYCLIC PHOSPHODIESTERASE (CPDASE) (BY SIMILARITY).
CC -1- SIMILARITY: SLIGHT SIMILARITY OF THE ACTIVE SITE REGION OF THE
CC LIGASE DOMAIN WITH t4 RNA-LIGASE ACTIVE SITE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L13380; AAA34373.2;
KW tRNA processing; Ligase; Hydrolyase; Endonuclease; Kinase; Transferase;
KM Multifunctional enzyme; ATP-binding.
FT BINDING 108 108 AMP (BY SIMILARITY);
SQ SEQUENCE 832 AA; 95569 MW; F0C801F633240130 CRC64;

Query Match 4.2%; Score 100; DB 1; Length 832;
Best Local Similarity 20.0%; Pred. No. 4.8;
Matches 72; Conservative 49; Mismatches 93; Indels 146; Gaps 17;

QY 77 DGAVALGAILNLEVKKKKREP-----THTLSLTRDCEHFKAEKKTQPLSKREVE 130
DB 466 DDOGVVLFDRNNSASRRRIETTDOKRDEH-LDDTVDKYIAI--NFTPEDLSBEL- 521
QY 131 FPIAYSMVNI-----HEKI--ENFERLIRAV-----YAPONT-----YCVHV-- 164
DB 522 WDTYRNVIORGNDHOSIKKSDENLVESVMKGFIOQYQIPINRSRDPDFVHILKLS 561
QY 165 -DEKSEPTFEAVKAIL-----SCFPN-----VFASKLVRYV 196
DB 582 KDNSSKSSLENNRIIIDLVDQNPDLIKEKPADELINECFOKALDYKPFYVKNMNTANT 641
QY 197 -----YAS-----WSRYQADLNCMEDLLQSS-----VPM 220
DB 642 KKPPTTYGIAHMYSSILENLEIVSHNEHFNONIKSHIQTEFHTVLGHASSKQDKAGRVK 701
QY 221 KYPLNTGGFDEPIKSNAM-----VOALMKMLNRRNMSSEVPP 258
DB 702 KKLIVKILGKDPKPKRSALKFADVKLLQIVITDKLACKVELIKYIDNDVLOSIEP 761
QY 259 KKEETRMKYHEEVVVDTLHLTNKKKDDPPYNTLMTGMAIVASRDFVQHVHLKNSQOL 318
DB 762 INKO-----LHTTICIPP-----ATAVESNMNLTLELYNDPDEQL 797
RESULT 11

```

```

TNPA_BACTU
ID TNPA_BACTU STANDARD: PRT: 987 AA.
AC P10021;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposase for transposon Tn4430.
GN TNPA.
OS Bacillus thuringiensis.
OG Plasmid pg12.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1428;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H1.1;
RX MEDLINE=89098342; PubMed=3211758;
RA Mahillon J., Seurinck J.;
RT "Complete nucleotide sequence of pg12, a Bacillus thuringiensis
RT plasmid containing tn4430."
RT Nucleic Acids Res. 16:11827-11827(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H1.1;
RX MEDLINE=88312602; PubMed=2842151;
RA Mahillon J., Ierectus D.;
RT "Structural and functional analysis of Tn4430: identification of an
RT integrase-like protein involved in the co-integrate-resolution
RT process."
RT EMBO J. 7:1515-1526(1988).
RT -1- FUNCTION: REQUIRED FOR TRANSPOSITION OF TRANSPOSON Tn4430.
RT -1- SIMILARITY: BELONGS TO THE TRANSPOSAE FAMILY 7.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X13481; CAA31833.1;
DR EMBL: X07651; CAA30492.1;
DR PIR: S00570; T0BS30.
DR InterPro: IPR002513; Transposase_7.
DR Pfam: PF01526; Transposase_7; 1.
KW Transposable element; Transposition; DNA-binding; DNA recombination;
KW plasmid.
SQ SEQUENCE 987 AA; 113644 MW; 919A64CD1D254F08 CRC64;

Query Match 4.1%; Score 97; DB 1; Length 987;
Best Local Similarity 19.5%; Pred. No. 10;
Matches 97; Conservative 63; Mismatches 172; Indels 166; Gaps 22;

QY 42 LESRESOSQYCRNIIYNFLK-----PAKRSLNCGVTRGDOEAVLQAI--- 85
DB 360 VESVEEAKOLSRPLNDYIDLNTFRSYRVAPTLRLSHFRATKSG--EPVLAQIDLRT 417
QY 86 --LNNEVKKKKREP-----TDHTLSLTRDCE----- 111
DB 418 HELN--ETGRRKRVHGAPLHFVSNRQKHVYDDGGINRRYELALTELNRHIRSGDIF 475
QY 112 -----HFKAERKFTQPLSKEE-----VEPIAYSMVTHEKI-----EN 145
DB 476 VSSGRHHA---FDDVLIPIDEMNEVSNIPNGLTAPLAKAEDYITDINRLNHELEWLSKN 532
QY 146 FERLLRAVYAPQNIYCVHDEKSPETFEAVKAIISCFPNVFLASKLVRYVYASMSR--- 202
DB 533 SEKLEGVDISQGLHVERLDROGTPEBAKAFSKLHSMLEPRIKITDILLIEV--ASWTGFHD 590
QY 203 --VOADLNCMEDLLQSSVPMKYTYLNTCGIDFPIKSNAMVQALMKMLNRRNMSSEVPPKH 260

```

```

Db 591 OFIHASTNOSPDEONIVLA-TLMAMGTINIGLTKMAENTPGISYROMANA----- 640
OY 261 KETRMK-YHFEVVRDTLHNTKKKDDPPNYLMTFGMATIVASRPVQ-----HYLK 311
Db 641 ---SQRMTDDANVRKOSTILYNFQKE---OKLSYWGDC-TTSSSDMRSLIVRSIHADS 694
OY 312 ND-----KSQOLIEWKDTYSPDEHLMTLQARAWMPGSPVNHPKYDISDMSIAR---- 362
Db 695 NPHYCGKGGCTIYREFVSDQLS-----AVHVAVITTNARDALH 731
OY 363 -LYKMOGHEGDDIDKGAIPAPCSGTHORALICYGAGDLNMLNHLANKPDKYDDNAL 421
Db 732 VLDGLHHTETDKIEHYHTDAGTYDOYFAL-----THLGFREFAPRIKDLAD 779
OY 422 QCL-----EYLYRKAI 433
Db 780 TKLFSTPGGEYENVAL 797

RESULT 12
MSS4_YEAST
ID MSS4_YEAST STANDARD: PRT: 779 AA.
AC P38994;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable phosphatidylinositol-4-phosphate 5-kinase MSS4 (EC 2.7.1.68)
DE (1-phosphatidylinositol-4-phosphate kinase) (PIPSK) (PtdIns(4)P-5-
DE kinase) (Dlphosphoinositide kinase).
GN MSS4 OR YDR208W OR YD8142A.05.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94203175; PubMed=8152413;
RA Yoshida S., Ohya Y., Nakano A., Anraku Y.;
RT "Genetic interactions among genes involved in the STT4-PKC1 pathway
RT of Saccharomyces cerevisiae."
RL Mol. Genet. 242:631-640(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Oliver K., Shore L., Harris D., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
RL -i- FUNCTION: CATALYZES THE PHOSPHORYLATION OF PHOSPHATIDYLINOSITOL-4-
RL PHOSPHATE ON THE FIFTH HYDROXYL OF THE MYO-INOSITOL RING, TO FORM
RL PHOSPHATIDYLINOSITOL-4,5-BIPHOSPHATE (BY SIMILARITY).
CC -i- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol 4-
CC monophosphate = ADP + 1-phosphatidyl-1D-myo-inositol 4,5-
CC bisphosphate.
CC -i- PATHWAY: ACTS DOWNSTREAM OF STT4, BUT IN A PATHWAY THAT DOES NOT
CC INVOLVE PKC1.
CC -i- SIMILARITY: BELONGS TO THE PTDINS(4)P-5-KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D13716; BAA02869.1; -
CC DR EMBL: Z68194; CAA92347.1; -
CC DR PIR: S43271; S43271.
CC DR SGD: S0002616; MSS4.
CC DR InterPro: IPR002498; PIP5K.
CC DR Pfam: PF01504; PIP5K; 1.
CC DR SMART: SM00330; PIPKc; 1.

```

```

KW Transferase; Kinase.
FT CONFLICT 610 610 F -> L (IN REF. 1).
SQ SEQUENCE 779 AA; 89320 MW; F66B76622C9F8F6 CRC64;

Query Match 4.1%; Score 96.5; DB 1; Length 779;
Best Local Similarity 18.3%; Pred. No. 8.2;
Matches 76; Conservative 60; Mismatches 155; Indels 125; Gaps 16;

OY 28 KLSFLKCDSDHGLGLESRSOSQYCN---ILYFLKLPARKSINSQVTRDQEAVALDA 84
Db 293 KSTSLPPENDHYHQSRTGNHNAANAANNNTTTTAAGLKRSESAIAEIKMKROS 352
OY 85 ILNNLEVKRRREPFDTHYLSLRDCEHFKARKKFIQFLEKEVEFPYASVVIHEKIE 144
Db 353 LHKREMKRRKRTF-----LVDD-----DRVLGNKVSCHGVNFIAYNMTCIRVA 399
OY 145 -----NEERLLRAYV-----APQNIYCVHDEKSPETFEKAVKAIISCF 183
Db 400 VERCSGIMKPLTPADRFRTKLAFDYHGNELPSSQYAFKFDYCEVERE-LRALFGID 458
OY 184 PNVFLASKLVRYVYASMSKVQADLNC---MEDLLQSSVPMKFFLNCGDFPIKSAENV 240
Db 459 PADYLVLSRSXYL-----SELNSPGKSGSEFYYSRDYKIYKI----- 498
OY 241 QALKMLNGRNSMESEVPPKHKETRMKYHFEVVRDTLHNTKKKDDPPNYLMTFGNAYIV 300
Db 499 -----HHSE-----HHLKKHIOEYNNHRDNTLQCFYGLHRYK 535
OY 301 ASRDFQVHLKNPKSQOLIEWKDTYSP--DEHLMTLQARAWMPGSPVNHPRYDISDMT 358
Db 536 MPISF-QNKIKRRIYFLV--NMNLFPPLDHIHTYDLKGSFW--GRFTNLKERLAKDR 590
OY 359 STARLYKMOGHEGDDIDKGAIPAPCSGTHORALICYGAGDLNMLNHLANKPDR 414
Db 591 STRPVAK-----DLNMLEBQKI--KFGP 612

RESULT 13
UL87_HSVSA
ID UL87_HSVSA STANDARD: PRT: 731 AA.
AC Q01007;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Gene 24 protein.
GN 24.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wiltmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RL "Primary structure of the herpesvirus saimiri genome."
RL J. Virol. 66:5047-5058(1992).
CC -i- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER EBV BCRF1,
CC HSV-6 U58, HVS-1 24 AND HCMV UL87.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X64346; CAA45647.1; -
CC DR PIR: D36808; D36808.
CC DR InterPro: IPR004285; UL87.

```

DR Pfam: PF03043; UL87; 1.
SQ SEQUENCE 731 AA; 83572 MW; B6B35F2A8EBB3D3 CRC64;

Query Match 4.0%; Score 94.5; DB 1; Length 731;
Best Local Similarity 22.2%; Pred. No. 11;
Matches 81; Conservative 43; Mismatches 130; Indels 111; Gaps 21;

OY 6 RLCOHLVYMLAGCYMLATVATLALKLSF-RLKCDSDHDLGLESRESOSQYCRNLYNFKLPA 64
DB 24 RDCNENIQLKIFCCIALANQVSSLDENQDLCIGH---ESTFTICRAVRRLLLGSNNYPF 80
OY 65 KRSLNCSGV-TRGDQAVLAAILNLELVKKRREPTDTHYLSLTRDCEHFKAEKRTIQFP 123
DB 81 IDTLNESATGTRGPMYNGPFLIINN-----TDSFY-KLTLNICSN-----KYLPIV 124
OY 124 LSKEVEFPPIAYSMVTHEKEIENFERLLRAYAPQNIYCVHDEKSPTEFKAVKALISCF 183
DB 125 YSLETTDMP-----HEPL-----AYRAIFYPDL-----EQTPIDYCMFRIICR-- 163
OY 184 PNVEIASKLVRYVYASMSRVOADLNCMEDLLQSSVPMKYFLNTGCTDEPIKSNAEVQAL 243
DB 164 -----YVTHSELE---ECYEYFLATVSP---PVTNICK-----KNYLRVYSAL 200
OY 244 KMLNGRSMSESEVPPKHKETRMKYHFEVVR-----DTLHLTKKK-----DPPYNL 290
DB 201 KTLIPS-TVLATPNPPDQLEF---FKFSILSPQEWMSLNLTKTKKILAAVHSHPHIV 256
OY 291 TMTFGNMY-----IVASRDFVOHVLN-----PKSQL-----IEMVK 323
DB 257 KLCSONAFKEIKITDANFIEMQOTVNHVMPNFAETKRDGSHPLKVSLLPDGVKVV- 315
OY 324 DTYSF 328
DB 316 -IYFP 319

RESULT 14
THBG_BOVIN STANDARD; PRT; 411 AA.

ID THBG_BOVIN
AC 09T36;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thyoixine-binding globulin precursor (T4-binding globulin).
GN SERPINAT7 OR TBG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Janssen O.E., Lahner H., Grasberger H., Spring S.A., Einspanier R.;
RT "Characterization and primary structures of bovine and porcine
thyoixine-binding globulin";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAJOR THYROID HORMONE TRANSPORT PROTEIN IN SERUM (BY
SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER, FOUND IN PLASMA.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF204928; AAF15301.1; -
CC HSSP: P01011; ZACH.

DR InterPro: IPR000295; Leuserpin2.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin.1.
DR PRINTS: PR00780; LEUSERPIN1.
DR SMART: SM00093; SERPIN.1.
DR PROSITE: PS00284; SERPIN.1.
KW Serpin; Glycoprotein; Plasma; Signal.
FT SIGNAL 1 15
FT CHAIN 16 411
FT CARBOHYD 34 34
FT CARBOHYD 97 97
FT CARBOHYD 163 163
FT CARBOHYD 251 251
SQ SEQUENCE 411 AA; 46017 MW; 8823B56DA8386AC2 CRC64;

Query Match 4.0%; Score 94; DB 1; Length 411;
Best Local Similarity 20.7%; Pred. No. 5.5;
Matches 74; Conservative 55; Mismatches 122; Indels 106; Gaps 19;

OY 53 RNILYNFLKLPARKSINGCVTRGDQAVLAAILNLELVKKRREPTDTHYLSLTRDCEH 112
DB 62 QNIFFSVSIPIAGLAMSILGACSTQYQILEGLFNL-----DTYVAETIQGQFQH 112
OY 113 FKAERKTIQPLSKEEVEFPPIAYSMVTHEKEIENFERLLRAYAPQNIYCVHDEKSPET 172
DB 113 LICS---LNPP---KKEELQGNALFLGKQKPLEKFLDDV---KNLY-----ETEVF 157
OY 173 KEAVKALISCFPNVFIASKLVRYVYASMSRVOAD-----LNCMEDLLQSSV----- 218
DB 158 S-----TDFSNVSAQOEIN-----SHVEQTKGKIVGLQIDKPNITVLYNLYCF 204
OY 219 -----PMKYFLNTGCTDEPIKSNAEVQALKMLNGRSMSESEVPPKHKETRMKYHFEV 272
DB 205 KAQMANKFDSKTEEGSEFLVDKTTY-----QVPMNHQMEQY-YH-LV 246
OY 273 RDTLHLTKKKRDKPPYVNLMTFTGNAYIVASRDFVOHVLNKPKSOQLIEWKDTYSPDE-H 331
DB 247 DTELNCTVILQMDYSKNALALF-----VL--PKEGOM-EWEVGMSSTKLK 288
OY 332 LMAITLQARAMPQSVPR---HPKYDISDMTSTARLYVWQHGEGIDKGAVPAPCSGI 385
DB 289 KWNRLKRGWVDFVFPFSISATYDLDGI-----LTKM-----GIQDAFADNDFSG 336

RESULT 15

RASO_PYRHO STANDARD; PRT; 879 AA.

ID RASO_PYRHO
AC 058687;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR PH0929.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamaezaki J., Kusuda N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC Rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.

```

CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -1- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP000004; BAA30025.1;
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR001687; ATP_GTP_A.
DR InterPro: IPR001238; RecF.
DR InterPro: IPR003405; SMC_C.
DR Pfam: PF00470; RecF; 1.
DR Pfam: PF02483; SMC_C; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (By SIMILARITY).
FT DOMAIN 141 744 COILED COIL (POTENTIAL).
SQ SEQUENCE 879 AA; 103673 MW; F04E30FD1B0CB29 CRC64;

Query Match 4.0%; Score 93.5; DB 1; Length 879;
Best Local Similarity 20.1%; Pred. No. 16; Indels 109; Gaps 17;
Matches 70; Conservative 62; Mismatches 107;

QY 19 YMLATVALKLSFRLKCDSDHGLRESOSQYCHNLLYNFLKLPARKSINCQVT 74
DB 84 YKVLDFARNVSYLKRDLGRMRHVTETSMESVSFIDRIIPYVF-----LNAIYVR 136
QY 75 RGDEAVLQA-----ILNNLEVKKKRPEPTDTHYLSLTRDCEHFKAERKFIQRP 124
DB 137 OGQIDAILSEDETDRDKIYKEILNLDLKEA-----YDNIGKIRKIKYKYSI 181
QY 125 SKEVEFPPIAVSMVHEKIEINPERLLRAVYAPQNTYCVHVEKSPETFEAVKATISCF 184
DB 182 EEKE-----KFTM-----KTENIEDLIR-----TQEKSDFTFVLNFI----- 212
QY 185 NVFIASKLVRYVYASMSHVDLNCMEDLQSSVPKTYFLNTCGTDPPIKSNAEWYQAL 244
DB 213 -----RNISSNLPRRLRELBGIEKVEKT-----LEATFNSITELKRLG 251
QY 245 MINGR-----NSMESEVPKPKHKEKTRWKYHFEVYRDTLHLTNKKKDDPPYNTLMTGN 296
DB 252 ELNGKKGRLERIRQLGEGIEKRRKS--KELEEVYKELPELEKKETE--YRRLIEFKD 306
QY 297 AYIVASRDFVQH--VLKNPKSQQLIEWVKDTPSPDEHMLATILQRRARM 342
DB 307 EYLAKRNLEKRLGILSN-RLQEVKRRKIKDAES-----KVARIIRWI 346

```

Search completed: September 26, 2002, 09:51:24
Job time: 373 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 09:44:16 ; Search time 61.88 seconds
(without alignments)
1224.495 Million cell updates/sec

Title: US-09-874-390-2
Perfect score: 2360
Sequence: 1 MYQMKRICOLHYLMALGCMY.....NALQCLEXYLRKAYITTEL 438

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 segs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rotent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rviro:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	2360	100.0	438	4	095395
2	1949	82.6	440	12	099CM3
3	1939	82.2	440	11	091K2
4	1728.5	73.2	398	11	09DBA3
5	1107.5	46.9	428	11	064165
6	868	36.8	453	4	09PI09
7	715	30.3	356	11	09D2A8
8	531.5	22.5	314	4	09HCY8
9	379.5	16.1	412	4	0962C3
10	379	16.1	472	5	020406
11	373	15.8	420	5	0962C5
12	371	15.7	445	5	019730
13	367	15.6	401	5	002314
14	360.5	15.3	436	5	019729
15	352	14.9	425	5	002234
16	349.5	14.8	402	5	002315

17	336.5	14.3	352	5	0962C1	0962C1 caenorhabdi
18	287.5	12.2	283	5	0962C4	0962C4 caenorhabdi
19	271	11.5	470	5	016776	016776 caenorhabdi
20	267	11.3	467	5	09U3H9	09U3H9 caenorhabdi
21	259.5	11.0	322	16	09A4H4	09A4H4 caulobacter
22	254.5	10.8	189	5	0962C6	0962C6 caenorhabdi
23	254	10.8	454	5	045866	045866 caenorhabdi
24	253.5	10.7	513	5	09TVPO	09TVPO caenorhabdi
25	253.5	10.7	876	5	09W034	09W034 drosophila
26	248.5	10.5	513	5	09XTM4	09XTM4 caenorhabdi
27	232	9.8	489	5	021796	021796 caenorhabdi
28	219	9.3	459	5	002309	002309 caenorhabdi
29	215	9.1	864	11	09EP10	09EP10 rattus norv
30	213.5	9.0	486	5	093643	093643 caenorhabdi
31	212.5	9.0	367	5	022481	022481 caenorhabdi
32	210.5	8.9	478	5	09XX12	09XX12 caenorhabdi
33	206	8.7	865	4	09H1B5	09H1B5 homo sapien
34	205.5	8.7	753	5	002312	002312 caenorhabdi
35	205	8.7	616	11	09EP10	09EP10 mus musculu
36	199	8.4	472	5	076634	076634 caenorhabdi
37	198.5	8.4	789	11	09EP11	09EP11 mus musculu
38	197	8.3	827	4	09H1B6	09H1B6 homo sapien
39	195	8.3	821	11	09EP11	09EP11 rattus norv
40	183	7.8	424	10	09LE60	09LE60 arabidopsis
41	183	7.8	424	10	094A75	094A75 arabidopsis
42	177	7.5	446	10	09LNN5	09LNN5 arabidopsis
43	176	7.5	408	10	09Z0Z7	09Z0Z7 arabidopsis
44	166	7.0	395	10	09C9A1	09C9A1 arabidopsis
45	164.5	7.0	447	10	09LR71	09LR71 arabidopsis

ALIGNMENTS

RESULT ID	1	095395	PRELIMINARY:	PRT:	438 AA.
AC	095395				
DT	01-MAY-1999	(TREMBLrel. 10, Created)			
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE (N-ACETYL)				
DE	TRANSFERRASE 3, MUCIN TYPE).				
GN	C2/4GNT.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RP	SEQUENCE FROM N.A.				
RP	TISSUE=BRIN:				
RX	MEDLINE=99115671; PubMed=9915862;				
RA	Yeh J.C., Ong E., Fukuda M.;				
RT	"Molecular cloning and expression of a novel beta-1, 6-N-				
RT	acetylglucosaminyltransferase that forms core 2, core 4, and I				
RT	branches.";				
RT	J Biol. Chem. 274:3215-3221(1999).				
RL	[2]				
RP	SEQUENCE FROM N.A.				
RP	MEDLINE=99143102; PubMed=9988682;				
RA	Schientek T., Nomoto M., Levery S.B., Merx G., van Kessel A.G.,				
RA	Bennett E.P., Hollingsworth M.A., Clausen H.;				
RT	"Control of O-glycan branch formation. Molecular cloning of human CDNA				
RT	encoding a novel beta1,6-N-acetylglucosaminyltransferase forming core				
RT	2 and core 4.";				
RT	J. Biol. Chem. 274:4504-4512(1999).				
RL	[3]				
RP	SEQUENCE FROM N.A.				
RP	TISSUE=COLON ADENOCARCINOMA;				
RA	Strausberg R.;				
RA	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF102542; AAD10824.1; -				
DR	EMBL; AF038650; AAD21525.1; -				

DR EMBL: BC017032; AAH17032.1; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 KW Transferrase; Glycosyltransferase.
 SQ SEQUENCE 438 AA; 50863 MW; 1F0A7B451C8407 CRC64;

Query Match 100.0%; Score 2360; DB 4; Length 438;
 Best Local Similarity 100.0%; Pred. No. 3.5e-189;
 Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MWOMKRLCOLHYLMALGCMYMLATVALKLSFRLKCDSDHGLGESRESOSQYCRNIIYNFL 60
 DB 1 MWOMKRLCOLHYLMALGCMYMLATVALKLSFRLKCDSDHGLGESRESOSQYCRNIIYNFL 60
 OY 61 KLPARKSINSGVTRGDQDAVLAQILNNLEVKKKREPFTDTHYLSLRDCEHFAERKFI 120
 DB 61 KLPARKSINSGVTRGDQDAVLAQILNNLEVKKKREPFTDTHYLSLRDCEHFAERKFI 120
 OY 121 QFPLSKEEVEEPPIAYSMVHEKIEENFERLLRAYAPONICYVHDEKSPETFEKAVKAI 180
 DB 121 QFPLSKEEVEEPPIAYSMVHEKIEENFERLLRAYAPONICYVHDEKSPETFEKAVKAI 180
 OY 181 SCFPNVFIASKLVRYVYASMSRYOADLNCMEDLLQSSVPKMYFLNTGCTDPPIKSNAMV 240
 DB 181 SCFPNVFIASKLVRYVYASMSRYOADLNCMEDLLQSSVPKMYFLNTGCTDPPIKSNAMV 240
 OY 241 QALKMILNGRNSMESEVPKHKETRMKYHFEVYRDTHLTNKKKDDPPRYNLMTFTGNAYIV 300
 DB 241 QALKMILNGRNSMESEVPKHKETRMKYHFEVYRDTHLTNKKKDDPPRYNLMTFTGNAYIV 300
 OY 301 ASRDEVOHYLKNPKSQQLIEWVKDYSPPDEHLMATLQARMPSPSVNHPKXDISDPTS 360
 DB 301 ASRDEVOHYLKNPKSQQLIEWVKDYSPPDEHLMATLQARMPSPSVNHPKXDISDPTS 360
 OY 361 ARLVKGQHGEGDIDGAPYAPCSGIHQRAICYVAGADLNMQLQNHLLANKFDPKVDNA 420
 DB 361 ARLVKGQHGEGDIDGAPYAPCSGIHQRAICYVAGADLNMQLQNHLLANKFDPKVDNA 420
 OY 421 LQCLEEYLRKKAITYGTEL 438
 DB 421 LQCLEEYLRKKAITYGTEL 438
 RESULT 2
 O99CWM3 PRELIMINARY; PRT; 440 AA.
 AC O99CWM3;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DE VIRAL BETA-1,6-N-ACETYLGUCOSAMINYLTRANSFERASE.
 OS Bovine herpesvirus 4.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae.
 OX NCBI_TaxID=10385;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20583805; Pubmed=1152491;
 RA Zimmerman W., Broll H., Ehlers B., Buhk H.-J., Rosenthal A.,
 RA Goltz M.;
 RT "Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and
 RT Identification of an Origin of DNA Replication";
 RL J. Virol. 75:1186-1194(2001).
 DR EMBL: AF318573; AK07999.1; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 KW Transferrase; Glycosyltransferase.
 SQ SEQUENCE 440 AA; 50763 MW; C88B3EAB8FDB8D9 CRC64;

Query Match 82.6%; Score 1949; DB 12; Length 440;
 Best Local Similarity 82.0%; Pred. No. 8.3e-155;

Matches 360; Conservative 38; Mismatches 39; Indels 2; Gaps 2;

OY 1 MWOM-KRLCOLHYLMALGCMYMLATVALKLSFRLKCDSDHGLGESRESOSQYCRNIIYNFL 59
 DB 3 MAGWKKKLCGHHLMALGCMYMLAVVSLRLSLRPFKCDVSDLESRDPOQHCRODMLYNS 62
 OY 60 LKLPARKSINSGVTRGDQDAVLAQILNNLEVKKKREPFTDTHYLSLRDCEHFAERKFI 119
 DB 63 LKLPARKSINSGITRDOQDAVLAQILNNLEVKKKREPFTDTHYLSLRDCEHFAERKFI 122
 OY 120 IOPPLSKEEVEEPPIAYSMVHEKIEENFERLLRAYAPONICYVHDEKSPETFEKAVKAI 179
 DB 123 IOPPLSKEEVEEPPIAYSMVHEKIEENFERLLRAYAPONICYVHDEKSPETFEKAVKAI 182
 OY 180 ISCFPNVFIASKLVRYVYASMSRYOADLNCMEDLLQSSVPKMYFLNTGCTDPPIKSNAMV 239
 DB 183 ISCFPNVFIASKLVRYVYASMSRYOADLNCMEDLLQSSVPKMYFLNTGCTDPPIKSNAMV 242
 OY 240 VOALKMILNGRNSMESEVPKHKETRMKYHFEVYRDTHLTNKKKDDPPRYNLMTFTGNAYI 299
 DB 243 VOALKMILNGRNSMESEVPKHKETRMKYHFEVYRDTHLTNKKKDDPPRYNLMTFTGNAYI 301
 OY 300 VASRDEVOHYLKNPKSQQLIEWVKDYSPPDEHLMATLQARMPSPSVNHPKXDISDPTS 359
 DB 302 VASRDEVOHYLKNPKSQQLIEWVKDYSPPDEHLMATLQARMPSPSVNHPKXDISDPTS 361
 OY 360 IARLVKGQHGEGDIDGAPYAPCSGIHQRAICYVAGADLNMQLQNHLLANKFDPKVDNA 419
 DB 362 IARLVKGQHGEGDIDGAPYAPCSGIHQRAICYVAGADLNMQLQNHLLANKFDPKVDNA 421
 OY 420 ALOCLEEYLRKKAITYGTEL 438
 DB 422 ALOCLEEYLRKKAITYGTEL 440

RESULT 3
 O91ZK2 PRELIMINARY; PRT; 440 AA.
 AC O91ZK2;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE BETA-1,6-N-ACETYLGUCOSAMINYLTRANSFERASE.
 GN BORF3-4.
 OS Bovine herpesvirus 4.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae.
 OX NCBI_TaxID=10385;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=V. TEST;
 RX MEDLINE=20283898; Pubmed=10811884;
 RA Vanderplassen A., Markine-Gorlaynoff N., Lomonte P., Suzuki M.,
 RA Hiraoka N., Yeh J.-C., Bureau F., Willems L., Thiry E., Fukuda M.,
 RA Pastoret P.-P.;
 RT "A multipotential beta-1,6-N-acetylglucosaminyl-transferase is encoded
 RT by Bovine herpesvirus type 4";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:5756-5761(2000).
 DR EMBL: AF231105; AAF72001.1; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 KW Transferrase; Glycosyltransferase.
 SQ SEQUENCE 440 AA; 50701 MW; 71B9C3B6BA4949BD CRC64;

Query Match 82.2%; Score 1939; DB 12; Length 440;
 Best Local Similarity 82.0%; Pred. No. 5.7e-154;
 Matches 360; Conservative 36; Mismatches 41; Indels 2; Gaps 2;

OY 1 MWOM-KRLCOLHYLMALGCMYMLATVALKLSFRLKCDSDHGLGESRESOSQYCRNIIYNFL 59
 DB 3 MAGWKKKLCGHHLMALGCMYMLAVVSLRLSLRPFKCDVSDLESRDPOQHCRODMLYNS 62

Qy	60	LKLPAAKSIINSQVTRGDQAVLOALINLNEVKKKKRREPFDHYLSLTFDCEHFKAKERF	119
Db	63	LKLPAAKSIINSQGITTRDQDAVQVALLDNLNEVKKRPPRLDTIYYLWITDCEHFKAKERF	122
Qy	120	IQPPLSKSEEEFPIAYSMVITHEKIEENFERLLRAVAPQNTYCVNHDEKSPETFEKAVKAI	179
Db	123	IQPPLSKSEEDLPIAYSMVYHEKIEENFERLLRAVAPQNTYCVNHVYDKSPETFEKAVKAI	182
Qy	180	ISCFPNPFIASKLYRVVYVYASRSKRVQADLNCMEDLLOSVPMKFLWTGCTDPPKISNAEM	239
Db	183	ISCFPNPFIASKLYRVVYVYASRSKRVQADLNCMEDLLOSVPMKFLWTGCTDPPKISNAEM	242
Qy	240	VQALMKLNGRSMSESEVPNPKHKEFETRIKREYVVDRLHLTNKKKDDPPNLMETGNATY	299
Db	243	VQALMKLNGRSMSESEVPNPKHKEFETRIKREYVVDRLHLTNKKKDDPPNLMETGNATY	301
Qy	300	VASDFEVOHVLKKNPKSQQLIEWKDYTSPEDEHLMATLQARARMPGSGVPHNPKYDIDSMTS	359
Db	302	VASDFEVOHVLKKNPKSQQLIEWKDYTSPEDEHLMATLQARARMPGSGVPHNPKYDIDSMTS	361
Qy	360	IARLVKMGHGGIDDKAPAPCSGTHORACVYGGADLNMWLONHHLILANFPDPRVDN	419
Db	362	IARLVKMGHGGIDDKAPAPCSGTHORACVYGGADLNMWLONHHLILANFPDPRVDN	421
Qy	420	ALQCLEEYLRKATYGTSL 438	
Db	422	VLOCLEEYLRKATYGTSL 440	
RESULT	4		
Q9DBA3			
AC	Q9DBA3;	PRELIMINARY;	PRT; 398 AA.
DT	01-JUN-2001 (Tremblrel. 17, Created)		
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)		
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)		
DE	2010013H22RIK PROTEIN.		
CN	2010013H22RIK.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;		
RX	MEDLINE=21085560; PubMed=11217851;		
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Aikawa T., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,		
RA	Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Sato R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Glassi C., King B., Kochiya H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,		
RA	Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA	Blake J., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,		
RA	Brownstein M.J., Bolt C., Fletcher C., Fujita P., de Bonaldo M.F.,		
RA	Guustlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA	Nordone P., Ring B., Schoenbach M., Rodriguez I., Sakamoto N.,		
RA	Sasaki H., Sato K., Rindwald C., Seya T., Shibata Y., Storch K.-P.,		
RA	Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilmink L.,		
RA	Wyrshak-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsuki S.,		
RA	Hayashizaki Y.;		
RT	"Functional annotation of a full-length mouse cDNA collection.";		
RL	Nature 409:685-690(2001).		
DR	EMBL; AK008234; BAB25548.1;		
DR	MGD; MGI:1919327; 2010013H22RIK.		
DR	InterPro: IPR003406; Branch.		
DR	Pfam; PF02485; Branch.1.		
SO	SEQUENCE 398 AA; 46000 MW; 275A8E93909CC4B6 CRC64;		

Query Match	73.2%;	Score 1728.5;	DB 11;	Length 358;
Best Local Similarity	77.9%;	Pred. No. 2e-136;		
Matches 311;	Conservative 44;	Mismatches 43;	Indels 1;	Gaps 1;

[illegible]

```

OY 186 VFASQLESVYASMSRVADLNCMEDLIQSSVPMKFFLNTGCTDPFPIKSNAMVOALKM 245
DB 176 VFASQLESVYASMSRVADLNCMEDLIQSSVPMKFFLNTGCTDPFPIKSNAMVOALKM 245
OY 246 LNCGRNMESEVPPKHKETFMKYHFEVYVDTLHLTNKKKDPDPNLMFTGNATVSRDP 305
DB 236 FPGENSELEFKMKPNKEERKKRYTVDDGLNTGVYKQAPRLKTLFSSSAFVYTR 295
OY 306 VOHLVKNPKSQQLIENVKDTYSPDEHLMATLQARMMPSVPMKPYDSDMTSIARLVK 365
DB 296 VGYVLENNKNIQKFMWAOPTYSDFELMATIQRIPEVPGSLPSSHKYDLSDMNAVARFVK 355
OY 366 MOGHEDIDKGAAPCSCGHORALCYGAGDMMMLQNHLLANKEDPRVDNALQCLE 425
DB 356 MOYFEBDSVNGADYPPCSGVHRSVCVFGVDLSMMLRKHFFPANKEDMDVDFALQCLE 415
OY 426 EYLYRAI 433
DB 416 EHLRHKAL 423

RESULT 6
OY 09P109 PRELIMINARY; PRT; 453 AA.
AC 09P109:
DT 01-OCT-2000 (TREMBLrel, 15, Created)
DT 01-OCT-2000 (TREMBLrel, 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel, 17, Last annotation update)
DE CORE 2 BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE 3.
GN C2GNT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20219156; PubMed=10753916;
RA Schwenker T., Yen J.C., Levery S.B., Keck B., Merx G.,
RA van Kessel A.G., Fukuda M., Clausen H.;
RT "Control of O-glycan branch formation. Molecular cloning and
RT characterization of a novel thymus-associated core 2 beta1,6-N-
RT acetylglucosaminyltransferase";
RT J. Biol. Chem. 275:11106-11113(2000).
DR EMBL: AF132035; AAF63156.1; -.
DR InterPro: IPR003406; Branch.
DR Pfam: PF02485; Branch; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 453 AA; 53052 MW; B4379AD4427F41CA CRC64;

Query Match 36.8%; Score 868; DB 4; Length 453;
Best Local Similarity 43.0%; Pred. No. 2,2e-64;
Matches 182; Conservative 71; Mismatches 146; Indels 24; Gaps 9;
OY 13 LNALGYMLATYALKSLRKLKCDSPHLGLESRSOSQICRNILYNFLKIPARSRINCSG 72
DB 23 LMLSLSLKLINL-----RRLFPQKDIYLYEYSLSTSPFRN--RTHVHXKEVREYVNCSCG 75
OY 73 VTGDDQAVLQATLNNLEVKKKR--EPPTDHYLSLTRDCDCHFAEKRFIOPLSKFEVER 131
DB 76 IV--EDEPL--ETGKSLERIRRDIIIDEDDDVYAMTSDCIIYTLGLYAKLVSKKEKSF 131
OY 132 PLAYSMVIEHEKIEFRLRAVYAPONIVCVHDEKSPETFEKAVKAIISCFPNVFIASK 191
DB 132 PLAYSLVYHKDAIMVERLLHAIYNQNIYCIHYDRKAPDFKVAAMNNKLACFSNIFASK 191
OY 192 LVAVYVYASMSRVADLNCMEDLIQSSVPMKFFLNTGCTDPFPIKSNAMVOALKM LGRNS 251
DB 192 LEAVEVAHISRLQADNLCSLDLSSKSIOMKYVYINLCGDFPLKSNELVSELKKGANM 251
OY 252 MESEVPPKHKETFMKYHFEVYV--DTLHL--TNKKKDPDPNLMFTGNATVSRDP 305
DB 252 LETVKKPNSKLEKFTYHHELRVYEVYVLPITINISKEAPRNIOIFGSAIFVLQAF 311

```

```

OY 306 VOHLVKNPKSQQLIENVKDTYSPDEHLMATLQARMMPSVPMKPYDSDMTSIARLVK 365
DB 312 VKYIFNNSIYODFFAWMSKIDYSPDEHFWATLIVPQIPGEI--SRSAQDVSDLSQKTRLVK 370
OY 366 MOGHEDIDKGAAPCSCGHORALCYGAGDMMMLQNHLLANKEDPRVDNALQCLE 425
DB 371 WNYEGGF-----YPSCTGSHLRVSCITYGAELRWILKQGHFANKFDSKVPDILIKCLA 425
OY 426 EYL 428
DB 426 EKL 428

RESULT 7
OY 09D2A8 PRELIMINARY; PRT; 356 AA.
AC 09D2A8:
DT 01-JUN-2001 (TREMBLrel, 17, Created)
DT 01-JUN-2001 (TREMBLrel, 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel, 17, Last annotation update)
DE 5330430K10RIK PROTEIN.
GN 5330430K10RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PIUTITARY GLAND;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana H.,
RA Saito T., Okazaki Y., Gojoubert T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleschmann W., Gaasterland T., Glissi C., King B., Kochava H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lustiglich S., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Lyons P., Marchionni L., Mashima J., Ringwald M., Rodriguez I., Sakamoto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyono-Oka K., Wang K.H., Wetz C., Whitaker C., Williams L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
DR EMBL: AK019924; BAB31918.1; -.
DR MGD: MGI:1925531; 5330430K10RIK.
DR InterPro: IPR003406; Branch.
DR Pfam: PF02485; Branch; 1.
SQ SEQUENCE 356 AA; 40446 MW; 0D67F48615656FD6 CRC64;

Query Match 30.9%; Score 715; DB 11; Length 356;
Best Local Similarity 46.9%; Pred. No. 9.9e-52;
Matches 136; Conservative 48; Mismatches 90; Indels 16; Gaps 4;
OY 100 DTHYLSLTR--DCEHFAEKRFIOPLSKFEVERPLAYSMVIEHEKIEFRLRAVYAPON 158
DB 61 DTSFPHLERISCQYRIQSHYIYISPLSEBAAPRLAYIWIHRDPTFERLFRAYIMPN 120
OY 159 IYCVHDEKSPETFEKAVKAIISCFPNVFIASKLVYVYASMSRVADLNCMEDLIQSSV 218
DB 121 YVCVHADSRAIDTFEKAVALRQLSCFPNFIASKLVYVYVGGFSRLQADLNCMDLAVSKY 180
OY 219 PMKYFLNTGCTDPFPIKSNAMVOALKM LGRNMESEVPPKHKETFMKYHFEVYV---- 275
DB 181 PMKYVLTGCGDFPLKTLNKEIINHLKRFYKGNITPGVLPAYIVRTKYVHQRKKG 240

```

QY 276 -LHITNKKKDPPEYNLTMTGNAYIVASRDVCHVKNPKKSQOLIEWYKDTYSPDEHMA 334
 DB 241 FMHKTNLKTPPRHQLIITGTATVALTRDFVNTILNDEKRIALILEMSKDTYSPDEHFW 300
 QY 335 TLOARMPGSPVPHPKYDSDMTSIARLVKMOGHEGDDIKGAPYAPCSG 384
 DB 301 TLNRIPGVPGSPMPNMSM-----TGNLRAYKW-----MDWEAKHGCGHG 339

RESULT 8
 Q9HCY8 PRELIMINARY; PRT; 314 AA.

AC 09HCY8;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-DEC-2001 (TREMBlrel. 16, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE D1113D9.2 (NOVEL PROTEIN SIMILAR TO BETA
 1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE) (FRAGMENT).
 GN D1113D9.2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Clark G.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL109806; CAC03741.1; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 FT NON_TER 1
 FT NON_TER 314
 SQ SEQUENCE 314 AA; 36042 MW; 8E1B3F3976B2E85 CRC64;

Query Match 22.5%; Score 531.5; DB 4; Length 314;
 Best Local Similarity 39.7%; Pred. No. 1.9e-36;
 Matches 100; Conservative 58; Mismatches 71; Indels 23; Gaps 6;

QY 123 PLSEVEEFPPIAYSMVTHEKIEFELRLRAYAPQNTYCVHDEKSEPTFEKAVKAIISC 182
 DB 1 PLSEBESDFSLAYITTHKELAMFVOLLRAIYVPONYVCHVDEKAPKRYTAVQTLVNC 60
 QY 183 FPNFVIFASKLVRYVYASMSRVQADLNCMEDLQSSVPKRYFLNFGTDFPKSNAEMVQA 242
 DB 61 FENVFISKTEKVAAYAGFTRQADINCKMAYVHKSQPMNYINLCGDFPKTKTNEIHHY 120
 QY 243 LK-MLNGRNSKSESVPRKH-KETRWKYHFEVY-RDTLHL--TNKKKDPPEYNLTMTGNA 297
 DB 121 IRSKMSKKNITPGVIOPLHISKTSQSHLEFVPGSIYAPPNPFKPPHNLITYGSA 180
 QY 298 YIVASRDFVCHVKNPKKSQOLIEWYKDTYSPDEHMAITLOARMPGSPVPHPKYDSDM 357
 DB 181 YVLTFRFVERILTIDIAKMDLQMSKDIRSEPHYVWTLNR--LKGPATP----- 228
 QY 358 TSIALRVKMOGH 369
 DB 229 -----AGWEGN 234

RESULT 9
 Q962C3 PRELIMINARY; PRT; 412 AA.
 AC 0962C3;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 GN GLY-18 (FRAGMENT).
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-N2.
 RA Warren C.E., Krizus A., Dennis J.W.;
 RT "Complementary Expression Patterns of Six Non-Essential Caenorhabditis
 elegans Core 2/1 N-Acetylglucosaminyltransferase Homologs."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY037797; AAK94762.1; -
 FT NON_TER 1
 SQ SEQUENCE 412 AA; 47583 MW; 5F9325D732DD4AAE CRC64;

Query Match 16.1%; Score 379.5; DB 5; Length 412;
 Best Local Similarity 25.5%; Pred. No. 1.4e-23;
 Matches 97; Conservative 76; Mismatches 150; Indels 57; Gaps 8;

QY 68 INCSGVTRGDEAVLQALNNLEVKRREPTDTHVYLSITRDCHEFAERKFFIOPISKE 127
 DB 34 INCLINFGSKNRQRLRVNSRSTEDK-----LLYSTRCQTLKSLPRFNKVPISPE 85
 QY 128 EVERPIAYSMVTHEKIEFELRLRAYAPQNTYCVHDEKSEPTFEKAVKAIISCPRVVF 187
 DB 86 EESFPLSYGLVYKELSGVLEMLSIYHPQNECYIAGVENSAPILFQMLKELSCFSGN 145
 QY 188 IASKLVRYVYASMSRVQADLNCMEDLQSSVPKRYFLNFGTDFPKSNAEMVQALKMLN 247
 DB 146 F-MKRPPLDWSHSLINSAICYCLEFSLKSDWRYFQYLSGVDIPLKTNLEWQILKLN 204
 QY 248 GRNMESESVPRKHKETRWKYHFEVYRDTLHITNKKKDPPEYNLTMTGNAYIVASRDVQ 307
 DB 205 GTANVEIK-PYQYORLNRK-----NETQSPLP---LFKSSLSLIPREAN 246
 QY 308 HVLNPKKSQOLIEWYKDTYSPDEHMAITLOARMPGSPVPHPKYDSDMTSIARLVKQ 367
 DB 247 HLSSSIPOQLLEFLRLNGADEGFWGLT-----FGKNKLFEDIGSLNFKEMI 294
 QY 368 GHEGDDIKGAPYAP-----CSGIHQAICVYAGDLMMLQNHILLAN 410
 DB 295 SYKNNVETNLTPPDGMYRTYSRQIMSKRCHMYMKAGSCVFGIDVPRLLKSLAVAH 354
 QY 411 KEDPRVDNALQC-LEEYLR 429
 DB 355 KEYLSEPEAYFCLKEHRR 374

RESULT 10
 Q20406 PRELIMINARY; PRT; 472 AA.
 AC Q20406;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE F44F4.6 PROTEIN (GLY-1).
 GN F44F4.6 OR GLY-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Coles L.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 investigating biology."
 RL Science 282:2012-2018(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-N2;

RA Warren C.E., Krizus A., Dennis J.W.;
 RT "Complementary Expression Patterns of Six Non-Essential *Caenorhabditis*
 RL *elegans* Core 2/I N-Acetylglucosaminyltransferase Homologs.";
 DR EMBL: 237092; CA85457.1; -;
 DR EMBL: AY037793; AAK94758.1; -;
 DR EMBL: AY037792; AAK94757.1; -;
 DR InterPro: IPR003406; Branch: 1.
 DR Pfam: PF02485; Branch: 1.
 SQ SEQUENCE 472 AA; 55412 MW; 36A1BD96D9C6245A CRC64;

Query Match 16.1%; Score 379; DB 5; Length 472;
 Best Local Similarity 25.3%; Pred. No. 1.8e-23;
 Matches 114; Conservative 83; Mismatches 162; Indels 92; Gaps 17;

QY 31 FRKCSOHLGLSERE-----SOSQYCRNLVNFKL-----PAKRS-----INCSG 73
 DB 13 FRKFECCLOGLMNVRFKIIISQFLYSFTYVYLRIGFVDYSLVFNKRLNIPLDCNSV 72
 QY 74 TRGDQRAVLQALINLNVKKRREPFDTHY-----LSLRDCEHFKAERKFIQPLSK 126
 DB 73 INGT-----NRKKISRARQ--WDWHPDWEHEIFNSRNCSTIDKVFNFTRIPISSR 122
 QY 127 EEEVEPIASWVTHEKTEFPERLRAVYAPQNTCYHNDKSEPTKEAVKAIISCPN- 185
 DB 123 EEAEPYPLAGLVYKTVVLTQMSLFYOPHFCITVDOSFNEKSVYQALPSCFPNN 182
 QY 186 -FEISKLVRYVYASWRYQADLNCMEDLQSSVPRKRYFNTCGTGFPIKSNAEWQAK 244
 DB 183 HVFIGP-----SOWGSGIILKNYTCFNNLSKSKQKMYQYLSGDLPIRTNEMVRIR 239
 QY 245 MNGRNSMESEVPKPKETRMKYHFEVVRDTLHLTNKKKDPPEYNTLMTGNAYIVASRD 304
 DB 240 ALNG--SNMTDYST-----FEVDR-----YKNMEGYLP--MPYVKSMSVVPPE 281
 QY 305 FVOHYLKNKRSQOLIMVADTYSPEHLWATLO-----RAR--WMPGSVPHN 349
 DB 282 GADYLLSSPRVOKLTKLSTKTPIDESFWSYSPALLPVGSIYRDLMLRKNFKLR 341
 QY 350 PKYDIDMTSIA-----RLVWQGHGDIKGAAPYAPCSGTHORATCYGAGDMMML 402
 DB 342 PPE-RTVNSIGTSTYIGRVQWGMQ-----KECFGVKDPSCVGEVDIEETM 388
 QY 403 QNHHLANKFDPKVDNALQCLEEYLRKAI 433
 DB 389 TRPELVAKLYLEFQPAFAFCMLKEVRRSL 419
 RESULT 11
 Q962C5 PRELIMINARY; PRT: 420 AA.
 AC Q962C5;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE GLY-16.
 GN GLY-16.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
 OC Rhabdilitidae; Peloderinae; *Caenorhabditis*.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=N2;
 RA Warren C.E., Krizus A., Dennis J.W.;
 RT "Complementary Expression Patterns of Six Non-Essential *Caenorhabditis*
 RL *elegans* Core 2/I N-Acetylglucosaminyltransferase Homologs.";
 DR EMBL: AY037793; AAK94757.1; -;
 DR EMBL: AY037792; AAK94756.1; -;
 SQ SEQUENCE 420 AA; 48325 MW; 6BDC5BE58711614 CRC64;

Query Match 15.8%; Score 373; DB 5; Length 420;
 Best Local Similarity 26.6%; Pred. No. 4.9e-23;
 Matches 105; Conservative 77; Mismatches 158; Indels 54; Gaps 11;

QY 57 YNFKLPARK-----SINCSGTRGDQRAVLQALINLNVKKRREPFDTHYLSLT 107
 DB 22 YLFVAKPSRRPPDLLDLEINCTNLOGLKNEKQLELNTKMTIEKLMNSDR----- 75
 QY 108 RCEHFKARKEFIQPLSKSEVEEFPYASWVTHEKTEFPERLRAVYAPQNTCYHNDK 167
 DB 76 -CHTVKSMFRFNEVPVLSSEEAFFPLSYGLVYKELSYQLFMLSIIYQONEXCIADVGN 133
 QY 168 SEPTKEAVKAIISCPNFIASKLVRYVYASWRYQADLNCMEDLQSSVPRKRYFNTNC 227
 DB 134 SASTFLLLEESDCEPNVHF--MKRPITWGSYEIINSYDICKFLSHKSNMKRYOYLS 192
 QY 228 GNDPEPIKSNAEWQAKMLNGRNSMESEVPKPKETRMKYHFEVVRDTLHLTNKKDPP 287
 DB 193 GVDIPKLTNEMVRKILKSLNGTANVAEIKYENR-----LLGONETESPLP 238
 QY 288 YNLTMTGNAYIVASRDFOHYLKNKRSQOLIMVADTYSPEHLWATLORAMM--PG 344
 DB 239 ----LFKSSLSLIPRKANAYLASSIPQOLLEFLNMTWADGEGVGLFGNKGLEDVPG 294
 QY 345 SVPNPKYDIDMTSTARKV-----QGHGDIKGAAPYAPCSGTHORATCYGAG 396
 DB 295 SL-NPEEWYTFKFNARANLTKPTDGFYIYSRQIWFESG-----CHNHKDGSCVFGIG 348
 QY 397 DLNMMLONHHLANKFDPKVDNALQCL-EEYLR 429
 DB 349 DVSNLQAKALVANKLYLTSEPAVCLLKEHR 382
 RESULT 12
 Q19730 PRELIMINARY; PRT: 445 AA.
 AC Q19730;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE F22D6.11 PROTEIN.
 GN F22D6.11.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
 OC Rhabdilitidae; Peloderinae; *Caenorhabditis*.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wilkenson U.;
 RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9069613; PubMed=9851916;
 RT "Genome sequence of the nematode *C. elegans*: A platform for
 RT investigating biology."
 RL Science 282:2012-2018 (1998).
 DR EMBL: Z71262; CA95817.1; -;
 DR InterPro: IPR003406; Branch: 1.
 DR Pfam: PF02485; Branch: 1.
 SQ SEQUENCE 445 AA; 51330 MW; 4380A86BAF05DC6D CRC64;

Query Match 15.7%; Score 371; DB 5; Length 445;
 Best Local Similarity 25.5%; Pred. No. 7.7e-23;
 Matches 98; Conservative 76; Mismatches 149; Indels 62; Gaps 9;

QY 68 INCSGTRGDQRAVLQALINLNVKKRREPFDTHYLSLRDCEHFKARKEFIQPLSK 127
 DB 62 INCLINFGSKNKNQRIYNSRSIEDK-----LLYSTDRQCLSLFRFNKVPVLSPE 113
 QY 128 EEEVEPIASWVTHEK-----IEFPERLRAVYAPQNTCYHNDKSEPTKEAVKAIISG 182

Db 114 EESPLSYGLLYKELSGVITQRYVLFMLSSITYPHONEYCIAVGENSAPITQNLKELSN 173
 QY 183 FPNVFIASKLVRYVYASMSRVOADLNCMEDLQSSVPKRYFLNCGDFPIKSAEMVOA 242
 Db 174 FSNHMF-MKRPIDMSGHEILINSAYDCLFSLHKSMDRRFQYLSIDPLKTLNEMVOI 232
 QY 243 LKMINGNSMSEVPPKHKETRMKYHFEVVRDLHLTKKKKDPYVNLMTGTNAVYAS 302
 Db 233 LKHLNGTANVEIK-PYQYQLRKG-----NETQSPLP-----LEKSSLSLIP 274
 QY 303 RDPVQVHLKLNKSOQLLEWVKDTPSPDEHLMATIQARMMGSPVNPBKIDISMTSIR 362
 Db 275 REANHLSSSSIPQOLLEFLNLTGIADEGFWGTL-----FGNNLFDIPGSLN 322
 QY 363 LVKMGHEGDDKCAPYAP-----CSGIRARICVYGAGDLMMMLONH 405
 Db 323 FKEMISTKNVETMLTPTDGMRYTISRDOIMSKPNCHNWKAKSCVFGIDGVRLLKSK 382
 QY 406 HLLANKFDPKVDNALOC-LEEYLR 429
 Db 383 ALVAHKFYLKSEPEAYFCLLEHRR 407

RESULT 13
 ID 002314 PRELIMINARY; PRT; 401 AA.
 AC 002314;
 DT 01-JUL-1997 (Tremblrel. 04, Created)
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE F15D6.2.2 PROTEIN.
 GN F15D6.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
 OC Rhabditidae; Pelodertinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Dobson R.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; Pubmed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL; Z83125; CAB05620.1; -.
 DR InterPro; IPR003406; Branch.
 DR Pfam; PF02485; Branch; 1.
 SQ SEQUENCE 401 AA; 45608 MW; EID3F8AD84C0873D CRC64;

Query Match 15.6%; Score 367; DB 5; Length 401;
 Best Local Similarity 25.5%; Pred. No. 1.5e-22;
 Matches 97; Conservative 74; Mismatches 143; Indels 66; Gaps 8;

QY 57 YNFKLAKR-----SINGSGVTRGDOEAVLQAILNLEVKKKRPTDTHYLSLT 107
 Db 22 YLFAKSSRPPTLDLDELTCNTINLQGLNKEKOLELINTKMEIDKLNSTDR----- 75
 QY 108 RDCHEFAERKFIQPLSKKEEFPPIASVYHKEIKENFEELRAVYAPQIYCVHDEK 167
 Db 76 ---CHTVSMFNFENVPLESEARPLSLGLVYKELSGVLEMLSSITQPONEYCIAVGEN 133
 QY 168 SPETFEAKVKAIIISCFPNVFIASKLVRYVYASMSRVOADLNCMEDLQSSVPKRYFLN 227
 Db 134 SASFTLLLEELSDCFN-----KRPIWGSYEILINSYDCLFSLHKSMMKRYFQYLS 188
 QY 228 GTDPIPSNAEMVQALMLNCRNSMSEVPPKHKETRMKYHFEVVRDLHLTKKKKDP 287
 Db 189 GVDIPLKTNLEMBRILKSLNGTANVEIKYENRR-----LQGNETESPLP 234

QY 288 YNLTMTGNAYIVASRDFVOHVLKPNKSOQLLEWVKDTPSPDEHLMATIQARMM---PG 344
 Db 235 ---LFKSSLSLIPRKAANYLASSIPQOLLEFLRMTWVADEGFWGTLFGNNKLFDPVG 290
 QY 345 SVRHPKRYDISMTSIRLVYVNOGHEDDIDGAYARCSGIRHQAICVYGAGDLMMMLON 404
 Db 291 S-----LNFEEHQIWFESG-----CHNHKQSCVFGIDGVSNLLOA 327
 QY 405 HLLANKFDPKVDNALOCL 424
 Db 328 KALVAHKLYLTSEPEAYFCL 347

RESULT 14
 ID 019729 PRELIMINARY; PRT; 436 AA.
 AC 019729;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE F22D6.12 OR GLY-19.
 GN F22D6.12 OR GLY-19.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
 OC Rhabditidae; Pelodertinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wilkinson J.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; Pubmed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=N2;
 RA Warren C.E., Krizus A., Dennis J.W.;
 RT "Complementary Expression Patterns of Six Non-Essential Caenorhabditis
 RT elegans Core 2/1 N-Acetylglucosaminyltransferase Homologs."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z71262; CAA95816.2; -.
 DR EMBL; AY037801; AAK94766.1; -.
 SQ SEQUENCE 436 AA; 50669 MW; 073ECE382E8B36F CRC64;

Query Match 15.3%; Score 360.5; DB 5; Length 436;
 Best Local Similarity 24.9%; Pred. No. 5.7e-22;
 Matches 102; Conservative 76; Mismatches 170; Indels 59; Gaps 9;

QY 39 HGLGLESRESQYCRNLIYNFLK-----LPAKRSINGSGVTRGDOEAVLQAILN 88
 Db 20 YFVESILFPRKQEDKNVSKQFLKSICTTASDSYLLDNMEIKSNILKGYK-----TN 71
 QY 89 LEVAKKREPTDTHYLSITRCEHFKAERKFIQPLSKKEEFPPIASVYHKEIKENFE 148
 Db 72 EKDIMHLDIEQLFSTNKCQFLTKLFRFNTPMSEAENHFLSGMLVYKDLPOVLF 131
 QY 149 LLRAVYAPQIYCVHDEKSPETFEKAVKALISCFPNVFIASKLVRYVYASMSRVOADLN 208
 Db 132 LLSSTIHPONEYCIAVGENSAPITQNLREYSTCFSNVH-MKRPPIWSGHEIIDSYYD 190
 QY 209 CMEDLQSSVPKRYFLNCGDFPIKSAEMVQALMLNCRNSMSEVPPKHKETRMKYH 268
 Db 191 CLEFLSHLETDRWFQYLSGVDIPLKTNLEWQILKHLNGTSNVEI---TNYQOAR--- 243
 QY 269 FEVVRDLHLTKKKKDPYVNLMTGTNAVYASRDPVQVHLKPNKSOQLLEWVKDTP 328
 Db 244 -----LTGKNENESP-LPLFKSSLSAIIIPRKAANOIASSTARKLLEFLMTEIA 292

OY 329 DEHLMATLQRAHMGSPVNPBPKYDISDWTSLARLVKMGHEGDI---DKGAPYAP--- 381
DB 293 DEGFNGTIL-----FGNKDQFNISGINSKDMMEYRDNONNININPTDGMSTYISRDQ 343
OY 382 -----CSCIHORATCYAGDLNMLNHLANKFDPKVDNALQCL 424
DB 344 IWPDELCKNMYMKDSCVFGIDVPRLRTSKALVAHFKYLSKSPFAEYFCL 392

RESULT 15

ID 002234 PRELIMINARY: PRT: 425 AA.
AC 002234;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE HYPOTHETICAL PROTEIN C54C8.11.
GN C54C8.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Dobson R.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; 283102; CAB05469.1; -
DR EMBL; AL032664; CAB05469.1; JOINED.
DR EMBL; AL032664; CAA21769.1; -
DR EMBL; 283102; CAA21769.1; JOINED.
DR InterPro; IPR003406; Branch.
DR Pfam; PF02485; Branch; 1.
KW Hypothetical protein.
SQ SEQUENCE 425 AA; 49137 MW; 71A78A4DD94AA079 CRC64;

Query Match 14.9%; Score 352; DB 5; Length 425;
Best Local Similarity 28.2%; Pred. No. 2.8e-21;
Matches 111; Conservative 58; Mismatches 167; Indels 58; Gaps 14;

OY 68 INCSGVTRGDQAVLQALINLNVKKRPPFTDTHLS--LTRDCEHFKAKRRTIOPLS 125
DB 41 LECQKMFDDDPVSLAGALFKFD---DREITLQIILKLSQENNAQCAEFQKITGFQEPIS 97
OY 126 KEEVEPRLAYSNVIEHEKLENFERLLRAYAPQNIYCVHDEKSPETFEKAVKAIISCPFN 185
DB 98 QELEPRLAYGMLVHGDFFVQLSLLSAIQPONQFCLAVDGNSSVEFGLVRLMSRCYGN 157
OY 186 V--FLASKLVRYVYASMSRVQADLNCMEDLLQSSVPWKYFLNTCGTFPIKSNAMVQAL 243
DB 158 IOYFTTDE--LRMGYEILTSVFOCVDYLAFLPSDMKTYFOYLSGVDAFLKSNLEMIRIL 214
OY 244 KMLNGRNSMESVPPKHKETFRWKYHFEVYRDTLHLNKKKDPPEYVNLMTETGNAYIVASR 303
DB 215 KALNG--SFNAELP-----FEYR-----LNKKR-PWSSPLPIKTSLSATFSR 256
OY 304 DFVQHLKNPKSQQLIEWVKDYSPDEHLMATLQ-----RARMPGSPVN---HPK 351
DB 257 KSNFVNSEKYLEQIDFLRGTTCADESIMATIGAPVKIKFSELPMPGFPDAKAWIHKN 316
OY 352 YDISDMTSLARLVKMGHEGDIKQ-----APYAPCSGIHORAICYGAGDLN 399
DB 317 Y---RRTGKLGKY-GENQKIDNGYYSRYQOYVNRAP-VKCKGYRYRLSCVFGYDLP 370
OY 400 WMLQNHHLANKFDPKVDNALQCLEEYLRKAI 433
DB 371 NLINRHELVAAHKLIFSYPAPAFPCLVENSRKSM 404

Search completed: September 26, 2002, 09:50:50
Job time: 394 sec

Db 361 ARLVKNQHGHDIDKGAFFAPCSGISHQRAICVYGAGDLNMLNHLANKPDKVDNA 420
QY 421 LQCELEYLRKAIYGTGL 438
Db 421 LQCELEYLRKAIYGTGL 438

RESULT 2
US-07-955-041-4
Sequence 4, Application US/07955041
Patent No. 5360733

GENERAL INFORMATION:

APPLICANT: FUKUDA, MINORU
TITLE OF INVENTION: A NOVEL BETAL-6
TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
TITLE OF INVENTION: LEUCOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
TITLE OF INVENTION: ENZYMAIC ACTIVITY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/955,041
FILING DATE: 19921001
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: CAMPBELL, CATRYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9294
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 428 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-955-041-4

Query Match 48.4%; Score 1142.5; DB 1; Length 428;
Best Local Similarity 57.5%; Pred. No. 6.3e-107;
Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;

QY 68 INCSGVRGQDAVLAQLNLEVK-KRREPFDTHLSLRDCEHKAERKFIQPLSK 126
Db 57 INCKVLQGVNLEOKYKLEILTVYFKKRRPWPDDYINNTSDCSSFIKRRKIYVEPLSK 116
QY 127 EEVEFFPIAYSMVHEKTEJENEBRLRAYAPONTYCVAVDEKSPETFEAVKAIISCPNV 186
Db 117 EEVEFFPIAYSIYVHHKTEMIDRLIRAIYMPONFCVAVDTKSEDSYLAAYMGIASCSFNV 176
QY 187 FIASKLRVYVYASWSRVQADLNCMEDLQSSVPKFFLNTCGTFPIKSNAEVQALKML 246
Db 177 FVASRLESVYVYASWSRVQADLNCMDKADYASAMWKYILNCGMDFPIKTNLEIYRKLL 236
QY 247 NGRNSMESEVPKPKHETRMKTHHEVVVDTLHLNKKKDPPEYILMTFTGNAYIVASDFV 306
Db 237 MGNNNLETERMPGSHKBRMKKRYEVVNGKLTNGVTMLPLLEPLSGSAFYVVSSEYV 296
QY 247 QHVLKPKSQQLLEWYKDYSPDEHLMATLQARAMPGSVPNHEKXIISDWTSTARLVKM 366

Db 297 GYVLQNKIKQKLEMMADTSPDEILMATQRIPEYVGSIPASHKYVLSMQAVAFVYM 356
QY 367 QGHGHDIDKGAFFAPCSGISHQRAICVYGAGDLNMLNHLANKPDKVDNALQCLEE 426
Db 357 QYFEGDVSKAGAPYPCGVAVRSVCIGAGDLNMLNHLANKPDKVDNALQCLEE 416
QY 427 YLRKAI 433
Db 417 HLRKAL 423

RESULT 3
US-08-227-455-4
Sequence 4, Application US/08227455
Patent No. 5624832

GENERAL INFORMATION:

APPLICANT: FUKUDA, MINORU
TITLE OF INVENTION: A NOVEL BETAL-6
TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
TITLE OF INVENTION: LEUCOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
TITLE OF INVENTION: ENZYMAIC ACTIVITY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/227,455
FILING DATE: 14-APR-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: CAMPBELL, CATRYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9957
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 428 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-227-455-4

Query Match 48.4%; Score 1142.5; DB 1; Length 428;
Best Local Similarity 57.5%; Pred. No. 6.3e-107;
Matches 211; Conservative 57; Mismatches 98; Indels 1; Gaps 1;

QY 68 INCSGVRGQDAVLAQLNLEVK-KRREPFDTHLSLRDCEHKAERKFIQPLSK 126
Db 57 INCKVLQGVNLEOKYKLEILTVYFKKRRPWPDDYINNTSDCSSFIKRRKIYVEPLSK 116
QY 127 EEVEFFPIAYSMVHEKTEJENEBRLRAYAPONTYCVAVDEKSPETFEAVKAIISCPNV 186
Db 117 EEVEFFPIAYSIYVHHKTEMIDRLIRAIYMPONFCVAVDTKSEDSYLAAYMGIASCSFNV 176
QY 187 FIASKLRVYVYASWSRVQADLNCMEDLQSSVPKFFLNTCGTFPIKSNAEVQALKML 246
Db 177 FVASRLESVYVYASWSRVQADLNCMDKADYASAMWKYILNCGMDFPIKTNLEIYRKLL 236
QY 247 NGRNSMESEVPKPKHETRMKTHHEVVVDTLHLNKKKDPPEYILMTFTGNAYIVASDFV 306

BEST AVAILABLE COPY

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2002, 08:15:46 : Search time 19.87 Seconds
(without alignments)
538.420 Million cell updates/sec

Title: US-09-874-390-2
Perfect score: 2360
Sequence: 1 MVQWRKLCQHLWALGCMY.....NALQCLBEYLRYKAIYGTGL 438

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

cal number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2.6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2.6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2.6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2.6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2.6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2360	100.0	438	4	US-09-233-506-2
2	1142.5	48.4	428	1	US-07-955-041-4
3	1142.5	48.4	428	1	US-08-227-453-4
4	1142.5	48.4	428	1	US-08-472-482-4
5	1142.5	48.4	428	1	US-08-487-069-4
6	1142.5	48.4	428	1	US-09-233-506-3
7	784	33.2	400	1	US-08-118-906-14
8	784	33.2	400	1	US-08-486-196-14
9	784	33.2	400	1	US-08-488-135-14
10	784	33.2	400	2	US-08-474-065-14
11	784	33.2	400	4	US-09-233-506-4
12	447	18.9	126	1	US-08-118-906-4
13	447	18.9	126	1	US-08-486-196-4
14	447	18.9	126	1	US-08-488-135-4
15	447	18.9	126	2	US-08-474-065-4
16	426	18.1	126	1	US-08-118-906-2
17	426	18.1	126	1	US-08-486-196-2
18	426	18.1	126	1	US-08-488-135-2
19	426	18.1	126	2	US-08-474-065-2
20	278	11.8	64	4	US-09-233-506-10
21	203.5	8.6	49	4	US-09-233-506-14
22	126	5.3	33	1	US-08-118-906-8
23	126	5.3	33	1	US-08-486-196-8
24	126	5.3	33	1	US-08-488-135-8
25	126	5.3	33	2	US-08-474-065-8
26	118	5.0	33	1	US-08-118-906-6
27	118	5.0	33	1	US-08-486-196-6

28	118	5.0	33	1	US-08-488-135-6	Sequence 6, Appl1
29	118	5.0	33	2	US-08-474-065-6	Sequence 6, Appl1
30	109	4.6	22	1	US-08-118-906-12	Sequence 12, Appl1
31	109	4.6	22	1	US-08-486-196-12	Sequence 12, Appl1
32	109	4.6	22	1	US-08-488-135-12	Sequence 12, Appl1
33	109	4.6	22	2	US-08-474-065-12	Sequence 12, Appl1
34	103	4.4	316	1	US-08-597-236-12	Sequence 12, Appl1
35	103	4.4	316	1	US-08-746-682A-12	Sequence 12, Appl1
36	101	4.3	788	2	US-08-907-166-6	Sequence 6, Appl1
37	100.5	4.3	2555	3	US-09-058-489-36	Sequence 36, Appl1
38	94.5	4.0	253	3	US-08-929-329-9	Sequence 9, Appl1
39	94.5	4.0	619	4	US-09-153-310-42	Sequence 42, Appl1
40	92.5	3.9	2289	3	US-09-051-019-2	Sequence 2, Appl1
41	91	3.9	3135	1	US-08-323-170B-2	Sequence 2, Appl1
42	91	3.9	3135	4	US-08-954-441-2	Sequence 2, Appl1
43	90.5	3.8	616	4	US-08-965-902A-2	Sequence 2, Appl1
44	90	3.8	1128	1	US-08-111-939-2	Sequence 2, Appl1
45	90	3.8	1194	4	US-09-092-508-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-09-233-506-2
; Sequence 2, Application US/09233506
; Patent No. 6136580
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Yeh, Jinn-Chern
; TITLE OF INVENTION: A Beta-1-6-N-Acetylglucosaminyltransferase That Forms
; FILE OF INVENTION: Core 2, Core 4 and I Branches
; FILE REFERENCE: P-LJ 3415
; CURRENT APPLICATION NUMBER: US/09/233,506
; CURRENT FILING DATE: 1999-01-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-233-506-2

Query Match	100.0%;	Score	2360;	DB	4;	Length	438;		
Best Local Similarity	100.0%;	Pred.	No. 6.9e-230;						
Matches	438;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	MVQWRKLCQHLWALGCMYLLATVAKLSFRLKCDSDHLGLSRESQSQCRLNLYNFL	60						
Db	1	MVQWRKLCQHLWALGCMYLLATVAKLSFRLKCDSDHLGLSRESQSQCRLNLYNFL	60						
QY	61	KLPKRSINCSGVTGQDQAVLQAILNNLEVKKKREPFTDTHYLSLTRDCEHFAEKRFI	120						
Db	61	KLPKRSINCSGVTGQDQAVLQAILNNLEVKKKREPFTDTHYLSLTRDCEHFAEKRFI	120						
QY	121	QPLSKEEVEFPFIAYSMVIEHKIENFPELLRAYAPQNIYCVHVDEKSPETFEKAVKAI	180						
Db	121	QPLSKEEVEFPFIAYSMVIEHKIENFPELLRAYAPQNIYCVHVDEKSPETFEKAVKAI	180						
QY	181	SCFPNVFIASKLVRYVYASWSRVQADLNCMEDLLQSSVPWKYFYLNTCGDTFPIKSNAMV	240						
Db	181	SCFPNVFIASKLVRYVYASWSRVQADLNCMEDLLQSSVPWKYFYLNTCGDTFPIKSNAMV	240						
QY	241	QALKMLNGRNSMESEVPFKHETRWKYHFEVVRDTHLNTNKKDPPPYNLMTFTGNATIV	300						
Db	241	QALKMLNGRNSMESEVPFKHETRWKYHFEVVRDTHLNTNKKDPPPYNLMTFTGNATIV	300						
QY	301	ASRDFYQHVHLKPNKSOQLIEWVKDITYSPDEHLWATLQARWMPGSPVHPKYDISDMSI	360						
Db	301	ASRDFYQHVHLKPNKSOQLIEWVKDITYSPDEHLWATLQARWMPGSPVHPKYDISDMSI	360						
QY	361	ARLVKQGHGEGIDKGPAYAPCSGIHQRAICVYVGAGDLNWLQNHLLANKFDPKVDDNA	420						